

From The Institution of Clinical Science and Education

Karolinska Institutet, Stockholm, Sweden

**THE NATIONAL PERINATAL PATIENT SAFETY PROGRAMME – THE CHALLENGES  
OF IMPLEMENTATION AND EVALUATION**

Charlotte Milde Luthander



**Karolinska  
Institutet**

Stockholm 2016

All previously published papers are reproduced with permission from the publisher.

Published by Karolinska Institutet.

Printed by AJ E-PRINT AB 2016

© Charlotte Millde Luthander, 2015

ISBN 978-91-7676-311-7

Cover picture: Stina Wirsén

*“A flower represents its seeds and its roots, as a person represents its path and its passions. Our prayer is that you nourish your path with your passion.”*

*Guru Singh in “A year of prayer”, through Sophia Göth, my guru.*



The national Perinatal Patient Safety Programme – the challenges of  
implementation and evaluation

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

**Charlotte Millde Luthander**

*Principal Supervisor:*

Associate Professor Charlotta Grunewald  
Karolinska Institutet  
Department of Women's and Children's Health  
Division of Obstetrics and Gynecology

*Opponent:*

Professor Lars Wallin  
Dalarna University  
School of Education, Health and Social studies

*Co-supervisor(s):*

Senior Lecturer Monica Nyström  
Karolinska Institutet and Umeå University  
Learning, Informatics, Management and Ethics  
(LIME), and Department of Public Health and  
Clinical Medicine

*Examination Board:*

Associate Professor Andreas Herbst  
Lund University  
Department of Obstetrics and Gynecology

Professor Ulf Högberg  
Uppsala University

Associate Professor Ulf Hanson  
Uppsala University  
Department of Department of Women's and  
Children's Health  
Division of Obstetrics and Gynecology

Department of Womens and Childrens Health  
Division of Obstetrics and Gynecology

Professor Mirjam Ekstedt  
Karolinska Institutet and Linnaeus University  
Department of Learning, Informatics,  
Management and Ethics (LIME) /System Safety  
in Health Care  
and Faculty of Health and Life Science



# 1 ABSTRACT

**Background:** Factors resulting in suboptimal care might be avoided by training and increased risk awareness. The Swedish Perinatal Patient Safety (PPS) programme was initiated in 2008 and a web-based CTG learning programme was launched in 2009. The objective of this thesis was to evaluate the impact and effects of a national programme to improve safety for the newborn by studying how the interventions affected professional teams, their knowledge in assessing cardiotocography, and by exploring local improvement efforts, as well as perinatal outcomes, before and after the main intervention i.e. the PPS programme.

**Material and method:** In **Paper I** semi-structured interviews explored how the core interventions affected the local teams and their mental models of patient safety improvement and readiness for change. In **Paper II** midwives and physicians classified one CTG tracing before and after getting access to the web-based learning programme in fetal surveillance. In **Paper III** midwives and physicians classified one CTG tracing individually and one pairwise. In **Paper IV**, multiple methods were used for data collection and analyses: a) the final reports from all 46 obstetric units were analysed, b) Apgar score <7 at 5 minutes (=asphyxia) before and after the PPS programme was assessed and c) quarterly cumulative incidence of newborns granted financial compensation from LÖF due to delivery related asphyxia during 2000-15 were analysed. **Paper V** was a criterion-based review of care processes during labour and delivery. Cases with asphyxia and controls with full Apgar were retrieved after the PPS programme and compared with data from a previous study.

**Results:** Self-assessment was a useful tool at the units and new team mental models about patient safety improvement emerged during the process. The peer review process was appreciated, but had no explicit connection to any measurable effects (**Paper I**). No significant improvement was seen after education (**Paper II**). There was no significant difference when CTG's were classified pairwise compared to individual classifications. (**Paper III**). In the final reports, 4/5 of units reported improved guidelines. New process measures and/or follow-up of outcome measures were reported in half of the units. The incidence of asphyxia remained unchanged. The incidence of settled claims showed a decreasing trend, 2012-14 (**Paper IV**). Supervision of fetal well-being was more often neglected after the PPS programme in both cases and controls. The odds for an overall risk of incautious management of oxytocin more than doubled, while the risk for traumatic instrumental delivery decreased significantly among cases (**Paper V**).

**Conclusions:** Extensive improvement efforts have been made in all obstetric units in Sweden. However, our results show that there are still gaps in delivery care that need improvement. When investing time and resources in multifaceted projects, it is crucial to

evaluate results in order to assess the benefits of the intervention and at the same time, understand facilitating and restricting features of implementation. Evaluation, however, is challenging in large-scale quality improvements with a diversity of actors interacting in complex environments that are constantly changing. To secure a variety of evidence-based process and outcome measures already when planning the project, improves the chances of successful evaluation. Then it will be possible to continuously follow-up adherence and measure reliability in care processes.



## 2 LIST OF SCIENTIFIC PAPERS

- I. Monica E Nyström, Anna Westerlund, Elisabet Höög, **Charlotte Milde Luthander**, Ulf Högberg, Charlotta Grunewald. Healthcare system intervention for prevention of birth injuries - process evaluation of self-assessment, peer review, feedback and agreement for change. BMC Health Services Research 2012;12:274, DOI: 10.1186/1472-6963-12-274
- II. **Charlotte Milde Luthander**, Ulf Högberg, Monica E Nyström, Hans Pettersson, Ingela Wiklund, Charlotta Grunewald. The impact of a computer assisted learning programme on the ability to interpret cardiotocography. A before and after study. Sexual & Reproductive Healthcare Volume 3, Issue 1, March 2012, Pages 37 – 41, DOI:10.1016/j.srhc.2011.10.001
- III. Hanna Åmark, **Charlotte Milde-Luthander**, Gunilla Ajne, Ulf Högberg, Hans Pettersson, Ingela Wiklund, Charlotta Grunewald. Single versus pairwise interpretation of cardiotocography, a comparative study from six Swedish maternity units, Sexual & Reproductive Healthcare, Volume 5, Issue 4, December 2014, Pages 195 – 198, DOI:10.1016/j.srhc.2014.05.004
- IV. **Charlotte Milde Luthander**, Karin Källen, Monica E Nyström, Ulf Högberg, Stellan Håkansson, Karin Pukk Härenstam, Charlotta Grunewald. Results from a National Perinatal Patient Safety Program in Sweden: The Challenge of Evaluation, Acta Obstet Gynecol Scand 2016; 2016; 95:596 – 603 DOI: 10.1111/aogs.12873
- V. **Charlotte Milde Luthander**, Hans Järnbert-Pettersson, Ulf Högberg, Sophie Berglund, Charlotta Grunewald. No reductions of suboptimal obstetric care as assessed by criterion-based review in a Swedish health region (submitted).

# CONTENTS

1	Abstract.....	1
2	List of Scientific papers .....	3
3	List of abbreviations.....	6
4	Background.....	7
4.1	Patient safety .....	8
4.1.1	Human error.....	9
4.1.2	System safety .....	9
4.1.3	Learning from accidents.....	11
4.1.4	Measuring safety.....	12
4.1.5	Resilience .....	13
4.2	Challenges in obstetric care .....	14
4.2.1	Fetal surveillance .....	15
4.2.2	Oxytocin.....	15
4.2.3	Team-work and communication .....	16
4.2.4	Clinical practice guidelines.....	16
4.2.5	Quality indicators in obstetric care.....	17
4.3	The National Perinatal Patient Safety programme .....	21
4.3.1	Background.....	21
4.3.2	Empirical setting.....	21
4.3.3	Swedish National Patient Insurance Company LÖF .....	22
4.3.4	Specific interventions used in the PPS programme .....	22
4.3.5	The self-assessment instrument .....	24
4.3.6	The peer reviewers .....	26
4.3.7	Visits to the departments, feedback reports and mutual agreement on improvement measures .....	26
4.3.8	Time frame of the Perinatal Patient Safety programme .....	27
5	Aims of the thesis .....	28
6	Material and methods .....	30
6.1	Paper I .....	31
6.1.1	Study population and design .....	31
6.2	Papers II and III .....	31
6.2.1	Creation of a ‘reference standard’ CTG pool .....	31
6.2.2	Paper II .....	32
6.2.3	Paper III .....	33
6.3	Paper IV .....	34

6.3.1	Study population and design .....	34
6.4	Paper V.....	35
6.4.1	Study population and design .....	35
6.5	Statistical analyses .....	38
6.5.1	Paper I.....	38
6.5.2	Paper II.....	38
6.5.3	Paper III.....	38
6.5.4	Paper IV .....	38
6.5.5	Paper V .....	39
6.6	Ethical approvals and considerations .....	40
7	Results.....	41
7.1	Main findings - Paper I .....	41
7.2	Main findings - Paper II .....	41
7.3	Main findings - Paper III .....	42
7.4	Main findings - Paper IV.....	42
7.4.1	Results of the content analysis .....	42
7.4.2	Apgar score .....	43
7.4.3	Settled claims .....	44
7.5	Main findings - Paper V.....	44
8	Discussion.....	45
8.1	The implementation of the National Perinatal Patient Safety Programme ...	45
8.1.1	The PPS programme in the light of implementation science .....	46
8.2	Clinical guidelines (to standardise or not to standardise – that is the question).....	59
8.3	Methodological discussion and considerations.....	62
8.3.1	Qualitative data .....	62
8.3.2	Quantitative data .....	64
9	Conclusions and clinical implications .....	69
9.1	Planning for future improvement interventions .....	70
10	Acknowledgements .....	72
11	Sammanfattning på svenska.....	75
11.1	Slutsatser och överväganden.....	77
12	References .....	78

### 3 LIST OF ABBREVIATIONS

AOI	Adverse Obstetric Index
BMI	Body Mass Index
CFIR	Consolidated Framework for Advancing Implementations Science
CMO	Chief Medical Officer
CP	Cerebral Palsy
CTG	Cardiotocography
CUSUM	Cumulative Sum control chart
FBS	Fetal Blood Sample
FIGO	International Federation of Gynecology and Obstetrics
GA	Gestational Age
HIE	Hypoxic-Ischemic Encephalopathy
LÖF	Swedish National Patient Insurance Company
MBR	Medical Birth Register
NICU	Neonatal Intensive Care Unit
PDSA	Plan-Do-Study-Act
PPS programme	Perinatal Patient Safety Programme
RCA	Root Cause Analysis
SFOG	Svensk Förening för Obstetrik och Gynekologi
TMM	Team Mental Models

## 4 BACKGROUND

- “Is it not obvious we just *have* to do better...!” my senior colleague summarised in the end of the day. With high aspirations for patient safety goals, the obstetric unit where I work has annual meetings scrutinising severe adverse events and near-misses during the previous year. Analyses, made by a multi-professional group of obstetric staff, are presented and all front-line staff is invited to comment and suggest improvements. The narratives are always emotionally striking. The conclusion made by my colleague about what seemed so immediately obvious in the root-cause analysis presented, made me wonder. What does it really mean to ‘do better’? How is it done? What are the ingredients in doing better? Are there other explanations than insufficient effort to do the best we can when things turn out wrong?

Shortly after, I was engaged to participate in an improvement project with the aim to decrease sphincter injuries at delivery. The project was successful and through the initiators of the project, Marion Lindh and Magna Andreen Sachs, I had the opportunity to participate in several international Patient Safety Conferences. This was the start of my journey into the world of patient safety.

During my half-time seminar I was asked the very inspiring question: “Why don’t clinicians just follow the guidelines?” This question is central as it forces us to examine the components of standardisation, adherence to clinical guidelines, and why professionals ‘cut corners’ to get the job done, and it has followed me ever since.

Parallel to my doctoral studies, I am working as an obstetrician. Meeting women and their partners’ during their most exciting time of life, when they are about to become parents, is a grace. It also means that you sometimes end up in difficult and even frightening situations. Being responsible for two lives, detecting deviations from normal and act in time can be stressful. There is comfort in the thought that we have come far in the development of well-functioning teams, in which all participants contribute with their special skills, sharing the goal and responsibility to solve the situation. The delight of training communication skills and realising that we can actually improve is a major excitement to me. These occasions are also the source of constant questioning of our work-about - what is the essence of improvement?

## 4.1 PATIENT SAFETY

The possibility of cure and the risk of harm are closely linked in the practice of medicine and the awareness of this potential conflict is as old as medicine itself. Some 150 years ago, many complications were considered unfortunate but inevitable side effects of medical interventions. Over the past decades there has been a comprehensive shift in the view of outcomes in healthcare. In 1937, the confusion of an anesthetic medication and an antiseptic fluid for external use caused the death of four patients attending care at Maria Hospital in Stockholm for trivial disorders. The investigation, revealing organisational deficiencies, led to a new proclamation demanding health facilities to immediately report harm emerging from care. Until today, this has been the basis for the Lex Maria report system in Sweden (SOSFS 2005:28). Cochrane suggested in 1972 that only healthcare interventions that had been shown to be effective in well designed evaluations should be offered to patients. His engagement to make randomised control trials form the basis of medical interventions eventually led to the founding of The Cochrane Collaboration in 1993. Brennan and Leape's report in 1991 on the incidence of adverse events showed that a large number of patients were injured or even killed due to mistakes in healthcare (1). Ten years later the report 'An organisation with a memory' suggested a systematic data collection of adverse events to be used for continuous analysis, with the focus on learning from errors (2).

The desire to understand and control negative events has directed interest towards other safety critical areas, such as aviation and nuclear industries. Many important insights on the psychology of error originate from the study of accidents in these industries. However, there are important differences between the aviation industry and healthcare that need acknowledgment (3). Airplanes and nuclear power plants are constructed, complicated systems that in a sense are ultimately describable if the necessary number of equations are set up to calculate the way they work (p. 214) (4). Pilots supervise processes that to a large extent are automatised and delivered according to the model 'few-to-many' (two pilots and a small cabin team to many passengers). This makes the work process relatively stable and foreseeable, and risks can to a large extent be handled with standardised safety functions.

Healthcare, on the other hand, is not a technically constructed system. It is better described as a complex system composed of the multitude of interactions and relationships between patients, caregivers, technical components, knowledge and skills that is impossible to fully describe or control. This makes the system vulnerable to brittleness from within. Healthcare is also delivered from few-to-one or one-to-one, making individual skillfulness crucial to care outcomes (5).

#### **4.1.1 Human error**

“Systems may fail, break down, or fail to function but only people make errors” Vincent argues in his book Patient Safety (p. 132) (6). To disentangle the many facets of error it might help to describe them from three perspectives: behaviour, outcome and psychology (although it is important to keep in mind that it is a simplification, carrying limitations). Classification schemes have been developed to describe errors in detail, making it possible to predict and understand different types of adverse events. Those schemes mostly focus on behaviour and outcome aspects of error. The psychological aspects can be further divided into errors of actions and errors of knowledge and planning. For instance, being distracted either from things happening in the room or by things going on in one’s mind when performing a difficult task can result in failure. Mistakes, on the other hand, are what happen when the right action is applied in the wrong situations (7), e.g. increasing oxytocin at the end of a delivery, despite pathological CTG. The intention - to deliver the infant faster - might be justified but increasing the oxytocin drip in this situation impairs gas exchange, deteriorating the situation further for the infant.

A particular category of error is represented by violation, acts that in contrast to other errors are deliberate deviations from the standard procedure or guideline. The reasons are often to ‘cut corners’ in order to save time or to get the job done when resources are strained (p.73) (8). In obstetrics, it’s evident that despite many errors, the result is almost always successful. At the other end of this spectrum is the human factor or the human hero. Human ability to accommodate to the changing environment, compensate for small mistakes, improvise to ensure that care is delivered as expected are some examples of what prevents accidents from happening. Human flexibility is the reason why things go right despite complex safety conditions.

#### **4.1.2 System safety**

The first story of human error is the most obvious explanation when a catastrophic event is revealed, for instance a complex delivery with a vacuum suction that ends in the death of an infant. One of the early accident models describes error from a linear causation perspective where the adverse event is the results of a chain of linked events (9). In this ‘individually based approach’ the origin of error can be found at the beginning of the chain and the main reason is ‘human error’, a category of behaviours that include inadequate attention, negligence and recklessness, that can all be eliminated (10). The person most obviously associated with an accident is judged ‘accident prone’ or a ‘bad apple’ and safety will improve if the individual is told to “do better”, re-train, or is even removed. The system will then return to its’, by nature, stable and safe state. Such a repressive, exclusive attitude is still prevalent in some healthcare organisations. Staff involved in

serious adverse events might even be immediately suspended, and investigation, reflection and learning processes are initiated later, if at all.

However, there is a second story pertaining to safety issues that demands we take into account a series of events that are influenced by the complexity of the working environment and the wider organisational context and that only a detailed analysis will reveal (11). In this system based approach the person in the front line eventually 'making the mistake' is viewed as an inheritor of the instability and lack of barriers in the system that in unpredictable ways may harm the patient (12). With this perspective, the obstetrician performing the exemplifying vacuum suction is part of a team in need of improvement.

A system based approach to safety in healthcare has gained strong support as it creates better conditions for learning from mistakes, mainly for two reasons. First, if an individual is to blame there is significant risk that an investigation of relevant factors that led to an error will end when that person is identified. This prevents further learning about possible barriers to avoiding similar adverse events in the future. Second, in a culture where blame is a possibility, the willingness to report misses and near-misses will be obstructed. In 2011, a new law was instituted in Sweden, that decrees healthcare providers take the main responsibility for investigation of adverse events, applying a system-based and more proactive approach (13, 14).

According to James Reasons 'Swiss Cheese model', accidents occur due to a combination of active and latent failures (12). The active failures, such as lack of attention to or violation of rules are made by front-line staff. The latent conditions, for example understaffing or inadequate training are built into the design of the system and changes over time as procedures are adapted to new circumstances. The accident occurs when weaknesses in both active and latent conditions coincide and a 'tunnel of opportunity' emerges through the layers of barriers (Figure 1).



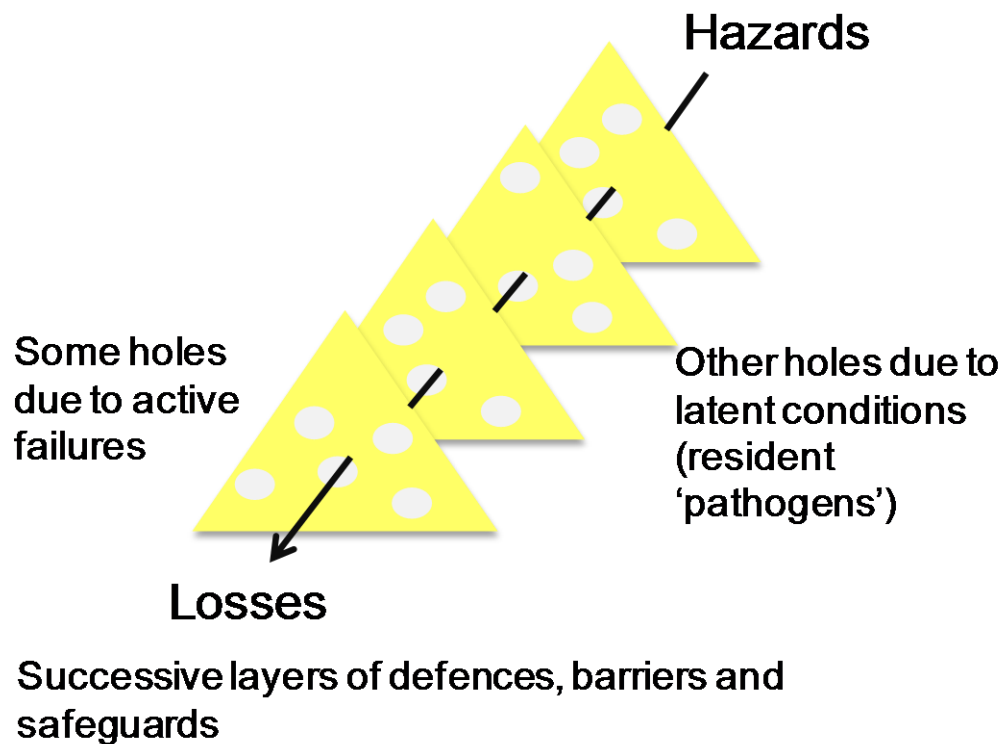


Figure 1. A visualisation of Reason's Swiss Cheese model

Many system based changes follows the rapid technical development introducing better computerised systems for medical journals, prompt availability of test answers, and computerised drug administration, among others. This development may improve safety but new techniques also introduce new risks within the system. At the same time, it's been suggested that the resilience of a system (defined by the ability to adjust operations to sudden changes and thereby continue to function and deliver safe care) is partly created by front-line staff constantly anticipating potential risks, discovering deviance and adapting to new circumstances (15). Weick and Sutcliff have called this ability 'mindfulness' (16) and the price for safety is 'chronic unease' (17).

#### **4.1.3 Learning from accidents**

The importance of organisational learning from adverse events is essential. There are a number of methods for trying to learn from looking back (event analysis). A widely used method in healthcare is called Root Cause Analysis (RCA) in which the most fundamental cause is sought out (18). It usually includes a team process where different perspectives are invited. However, there are limitations in the procedure from a system perspective as there is very rarely one single root cause to be found; it is something we construct. The construction of a cause is influenced by the investigators pre-conceptions and interpretations and can thereby be affected by many aspects such as politics, the struggle of available resources, and practical conditions of what is changeable (4). The linear

causation model is often implicit in the investigators understanding of how adverse events happen (19). As suggested by Hollnagel's principles 'What you look for is what you find' and 'What you find is what you fix), this too affects the result of the investigations and the actions suggested (20, 21).

Although looking into the driving mirror is necessary in order to capture important knowledge for the future, there are other pitfalls to keep in mind. We are affected by outcome bias, which is when knowledge of outcome influences the evaluation of the appropriateness of care delivered by others. Hindsight bias, on the other hand, is related to how we perceive the probability of an accident once we know the outcome.

"Hindsight ...knowledge means being able to look back, from the outside, on a sequence of events that led to an outcome that has already happened. It means that in hindsight people have almost unlimited access to the nature of the situation that surrounded people at the time...Hindsight allows investigators to pinpoint what people missed and should not have missed, what they did not do but should have done." Sidney Dekker, Patient Safety, p153 (4). In hindsight the complex reality of involved factors and uncertain outcomes runs the risk of becoming oversimplified, leading only to the inevitable outcome that we already know.

#### **4.1.4 Measuring safety**

Safety is generally perceived as the absence of unwanted events, a 'non-event'. This might explain why measuring and monitoring safety in healthcare is still a field of debate and evolution. Different approaches are needed depending on whether the focus is on error, harm, reliability, quality or culture.

Safety indicators can be viewed from two perspectives. Lagging indicators describe errors and adverse events that have already happened. Assuming there is a reliable reporting structure where staff feels comfortable with a non-punitive culture assuring that relevant data is collected, it is often possible to reveal patterns of risk.

Leading indicators, on the other hand are process measures indicating the built-in safety in a system. Through regular assessment of the organisational context such as adherence to guidelines, staffing and sensitivity to the conflict between economy and safety, threats can be identified and risks anticipated (8). In healthcare, leading indicators are still under development but proactive process measures might be used to reveal upcoming threats in the system.

#### 4.1.5 Resilience

The view of safety as a non-event is a perspective called Safety-I. Safety is defined by the absence of events. The aim of exploring the causes when things go wrong is to eliminate risk, improve barriers and bring the system back to its basically safe condition, where people 'work-as-imagined'. But focusing on the very rare events where 'human errors-causes' failure teaches us nothing about the many situations where human performance is the very reason for why things go right (Figure 2).

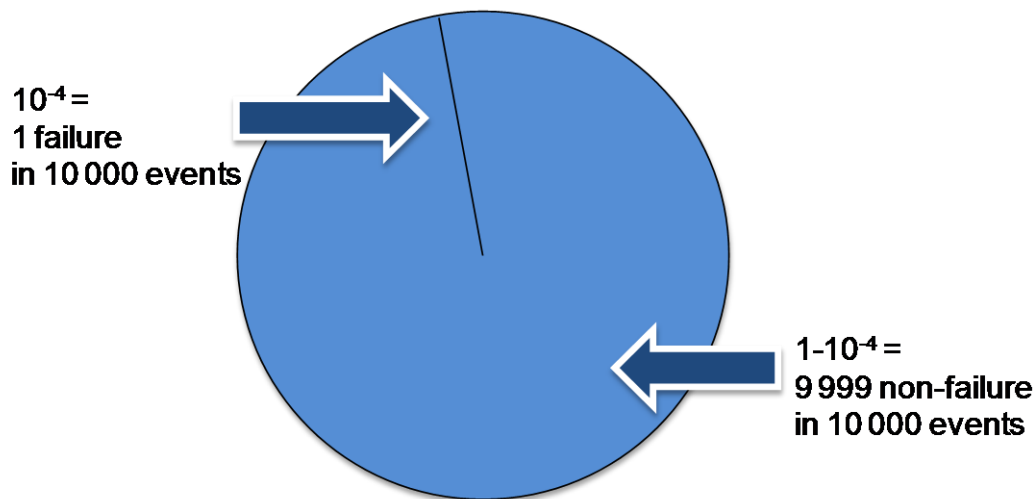


Figure 2. The imbalance between things that go right and things that go wrong (adapted after Hollnagel, Braithwaite and Wears (22))

Safety-II, on the other hand, explores the high frequency outcomes when things go right. The performance adjustments made by flexible humans compensating for lack of staffing, information, time and resources in the variable system is the basis for successful outcomes. By learning about these adjustments it is possible to understand why they sometimes fail. Both perspectives are needed in a resilient organisation.

## 4.2 CHALLENGES IN OBSTETRIC CARE

Both Safety I and Safety II perspectives are important to meet the challenges in perinatal care. There is a need to prevent injuries and act in time when required and, at the same time, avoid unnecessary interventions. Although severe delivery-related injuries to the neonate are rare in Sweden, and the neonatal mortality rate 0-27 days after birth is low (1.8 per 1000 live births in 2013), there is room for improvement. The exact number of children born with delivery related asphyxia is unknown but 20-30 children a year are granted economic compensation from the Swedish National Patient Insurance Company LÖF due to a delivery related asphyxia injury considered avoidable. These events are disastrous for the afflicted patients and their families. A previous study of children born in Stockholm in 2004-2006 with Apgar <7 at five minutes has shown that two thirds were subjected to suboptimal care during delivery, and so were one third of the healthy controls. It was concluded that some of these situations might have been avoided by education, training and increased awareness of risk areas (23).

Obstetric care is a highly complex and potentially hazardous activity and yet most situations will resolve normally, despite imperfect processes being identified when details are scrutinised. It has been calculated that one cerebral palsy and one perinatal death per 4,000 might be preventable based on figures in the 'Confidential inquiry into stillbirths and deaths in infancy', summarising that misinterpretation of CTG is still the most common cause of preventable errors with severe outcomes (24). This infrequency of severe outcomes makes obstetric care especially vulnerable to 'normalisation of deviance' (= slide to failure) where processes with hidden errors might continue for a long time without being discovered (25). Despite small, frequent errors in fetal surveillance made by the individual midwife or obstetrician, or in the team around the labouring woman, the risk of a severe outcome is low. This partly explains the overconfidence bias ("it never happened to me") and availability bias (it worked out fine the last time). Unless cumulative data on unusual events reveal the larger picture, the experience of the individual care giver will take precedence. The absence of severe outcomes is not equal to safe care.

Issues related to patient safety and in need of improvement within obstetric care are for example inconsistency in the interpretation of cardiotocography (CTG), as well as ambiguous expectations on actions aiming at improving the situation for the infant when pathological patterns occur (26-29). The management of oxytocin, communication and documentation gaps within the care giving team during delivery, and ambiguous guidelines are other examples as well as operative delivery procedures.

### 4.2.1 Fetal surveillance

Fetal assessment and monitoring during labour is one of the greatest challenges in current obstetrics (23, 26-28, 30, 31) (Figure 3). CTG is characterised by poor reproducibility (32), low specificity for signs of asphyxia, and high inter-and intra-observer variability (33-35). Adding fetal ECG ST-segment analysis (STAN) has been tried to improve specificity of the method but in a recent multicenter trial this did not improve perinatal outcomes (36). Yet, non-observance of signs of asphyxia has been highlighted and the need for persistent focus on this complicated analysis has been emphasised (23, 37, 38). There is still, after more than 30 years of routine use, no common language for CTG-patterns among front-line staff. This makes it very difficult to create common expectations for interventions when there are signs indicating that fetal well-being is endangered (29). A systematic review of 20 studies evaluating CTG education programmes showed that training resulted in increased CTG knowledge and interpretive skills, as well as higher inter-observer agreement and improved quality of care (39). To our knowledge it is still unknown whether a relatively high level of competence can be increased by education in a non-selected group of midwives and physicians. Other questions remain to be answered about the effective components in the above mentioned approaches, such as what methods for training are the most efficient, how often it should be repeated, and what other contextual factors are involved that affect the outcome of training (32).

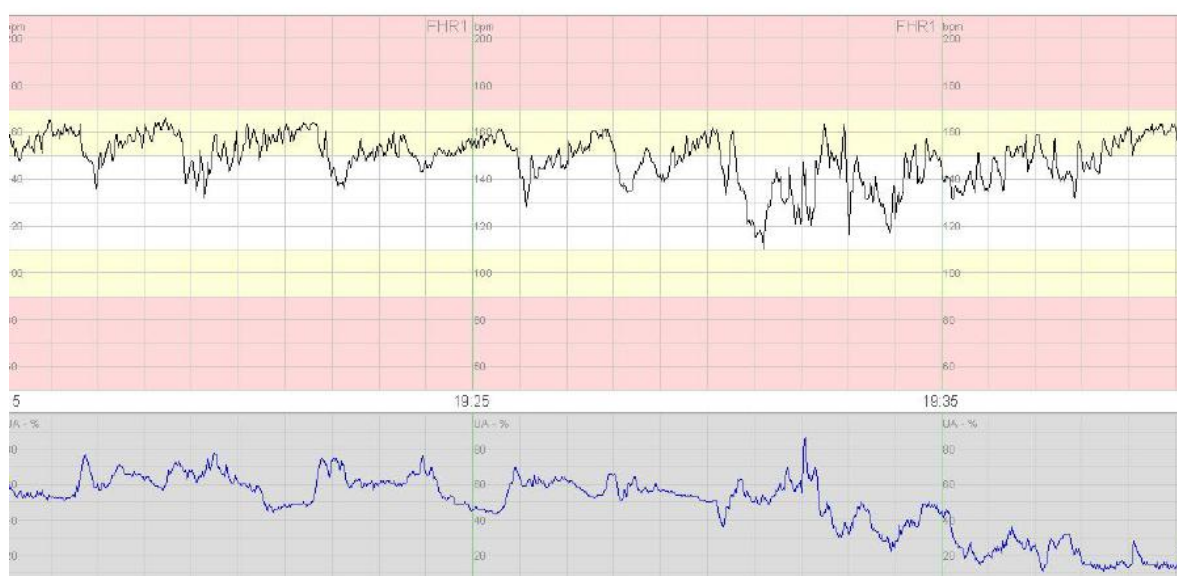


Figure 3. An intrapartum CTG registration.

### 4.2.2 Oxytocin

Oxytocin is a potentially dangerous drug. It is considered a high alert-medication by the Institute for Safe Medication Practices (40). In a study of 177 infants suffering from severe asphyxia at birth due to malpractice around labour, 71% were subjected to incautious use

of oxytocin (23). Acidemia at birth has been shown to be strongly associated with a hyperactive contraction pattern, most often due to oxytocin treatment (41). A more than three-fold risk of low Apgar score was seen if oxytocin was increased despite pathological CTG in a study of infants born with asphyxia in Stockholm 2004-2006 (37).

#### **4.2.3 Team-work and communication**

Dysfunctional teamwork is a well-recognised risk factor for adverse perinatal outcomes (26, 28, 42, 43). Over the past decades, obstetrics has gone through a major transition, from low risk operations controlled by midwives to high risk, monitored by midwives and obstetricians in teams. Although warnings about overly simplifying these complex situations and settling for the most obvious explanations are voiced (25, 44), a growing body of evidence highlights that simulation and communication training improves safety in obstetric settings (45, 46).

In escalating situations during delivery multi-professional teams involving midwives, obstetricians, anesthesiologist and neonatologists work together. These often complicated situations require that the team agrees on roles and functions and shares knowledge about the condition, possible chains of action, processes and procedures (47). The better the team can share these 'mental models' the better co-ordination and decision-making performance can be expected. Clear team mental models (TMMs) of evidence-based guidelines improves patient outcomes (48).

In some teams, experts contribute different pieces of information when building a common model. This is a strength when using multi-professional teams in improvement efforts. TMMs that promote learning and creative problem solving can be beneficial (49) in processes related to patient safety. Whether coherent TMMs can enhance change processes in patient safety work has not previously been described on a national level.

#### **4.2.4 Clinical practice guidelines**

Clinical guidelines are fundamental for high reliability care, but in many healthcare units there is a lack of unambiguous, easily accessible guidelines. The guidelines' function as helpful tools depends largely on how they are implemented and assimilated into the clinical practices. Passive dissemination (i.e. unplanned, uncontrolled) has very little effect on the behaviour of front-line staff whereas active, local implementation can increase compliance sustainably (50). This view on safety, presuming that a certain level of proceduralisation is necessary in order to protect patients from human fallibility, is contrasted by the approach whereby people create safety by individualising the application of guidelines to the specific situation and the unique individual they care for (6, 51). In practice, both perspectives are necessary. In Sweden, evidence-based recommendations on topics considered especially important are published on the National Society of Obstetricians and Gynecologists,

SFOG. They can be used to create locally adapted guidelines at the obstetric units but are not mandatory. Clinical practice guidelines and recommendations in the local settings have not been explored.

#### **4.2.5 Quality indicators in obstetric care**

In healthcare, an often used model for measuring safety includes measures of structure-, process- and outcome (52-54). Structure measures encompass the organisational context and include whether guidelines and key clinical protocols for, for example the management of oxytocin or fetal surveillance, are in place; if there is formalised interdisciplinary co-operation between, for example, neonatal, anesthesia and obstetric staff; the presence of a report system for adverse events; and systems for regular education in fetal surveillance for front-line staff. Most can be assessed by percentages or a yes/no answer, making them fairly easy to follow -up. However, questions about how an organisation can assess how well they learn from errors or how the safety climate changes with system improvement is more difficult.

Process measures estimate how care is delivered by, for instance, measures of adherence to an evidence-based guideline, i.e. if oxytocin guidelines are followed; if fetal surveillance is correctly and adequately documented in the medical charts; or if umbilical cord pH is taken and documented as prescribed. Front-line staff have been shown to prefer process measures as they are available for direct influence from changes made by staff themselves (54). There is also great potential to discover and learn from mishaps and miscommunications in all situations where an adverse event never became the ultimate outcome. Again, with a safety-II perspective, studying how health professionals bridge gaps in communication, anticipate conflicting goals, and manage the complex reality can give us new tools to measure safety.

Outcome measures traditionally report mortality and morbidity data. This is, however, not always the most suitable outcome, especially in settings where these severe adverse outcomes are very unusual. Attempts to join stakeholders and healthcare professionals in consensus around suitable outcomes have been made but need further effort.

In perinatal care there is an ongoing discussion on which outcome measures best represent quality of care and on how to reach consensus about these measures (55-59).

In Sweden the incidence of delivery related severe asphyxia is very low. There is, however, still no consensus on which outcome measures should be used and followed up or how this should be done. As within most healthcare settings around the world there is a lack of expertise in outcome and process analysis (60-62).

An easy to use, all else excluding measure of intrapartum asphyxia does not yet exist. Instead, a combination of indirect measures, such as Apgar, pH and lactate in the umbilical

cord, transfer to NICU, and development of hypoxic ischemic encephalopathy, among others, are used. Below are some examples of outcome measures often used in perinatal care presented in more detail: asphyxia, Apgar score, umbilical cord pH and lactate, AOI and CUSUM charts.

#### 4.2.5.1 Asphyxia

Asphyxia is a situation where the exchange of gases over the placenta is impaired. This leads to increasing levels of carbon dioxide and decreasing oxygen levels in circulating blood in the infant, eventually resulting in acidosis (63). During pregnancy this can be a slow, chronic situation due to a variety of maternal or neonatal complications, such as pre-eclampsia, diabetes or fetal anemia. Sudden detachments of the placenta from the uterus wall or a prolapse of the umbilical cord are situations where acute asphyxia evolves. The fetus uses compensatory mechanisms to minimise harm by reducing movement and increasing uptake of oxygen to fetal haemoglobin. If the hypoxia continues, it will precipitate a stress signal and the release of adrenalin and noradrenalin, resulting in a redistribution of blood from the peripheral parts of the body to the heart, brain and adrenal glands. Conditions crucial to how the infant reacts to asphyxia are gestational age, nutritional state, intrauterine infection and genetic prepositions.

During delivery virtually all neonates develop some degree of respiratory acidosis when carbon dioxide is not fully distributed from infant to mother, most often due to compression of the umbilical cord. If asphyxia has become more profound, different degrees of metabolic acidosis is seen. The cells in the tissue of the infant adapt to the lack of oxygen by using anaerobe metabolism with pH-lowering hydrogen ions and lactate as end products.

##### 4.2.5.1.1 Apgar

In 1953 the anesthesiologist Virginia Apgar presented a tool for rapid assessment of a newborn child to evaluate the need of immediate help to establish breathing. By evaluating heart rate, breathing, skin colour, muscle tone and reflex irritability, and grading them from 0-3, a sum from 0 to 10 is obtained. This is repeated at one, five and 10 minutes and when correctly used, it can be applied to assess fetal to neonatal transition and resuscitation progress. The advantage of using Apgar score when evaluating care is that it is registered at 99.6% of all deliveries reported to MBR in Sweden. Apgar scores 7 to 10 are considered normal; scores of 4 to -6 at five minutes may be the result of immaturity, maternal medication, or congenital malformations, as well as intrapartal asphyxia. The incidence of Apgar <7 at five minutes 1988-1997 was 0.76% in 1 million term newborns without malformations in Sweden (30) and the National Vital Statistics report an incidence of 1.5 % of all newborns in USA, 2005 (64).



Apgar < 4 is correlated to HIE, neonatal death and pathological CTG (65), is considered reliable as an indicator for intrapartum asphyxia and is also used as an indicator of quality in obstetric care (26, 56, 57, 59, 65, 66). In term infants an Apgar score of 3 at five minutes was associated with high risk for cerebral palsy (CP). Also an Apgar score of 6 at five minutes is associated with increased risk for CP (67), and Moster et al found a strong association between Apgar <7 at five minutes, neonatal death and CP (68). A score of 0-3 may correlate to neonatal mortality but is not alone usable as a predictor for later neurologic impairment (69). The predictive value for future impairment in newborns is questioned but when studying the academic achievements at 16 years of age, Apgar <7 at five minutes was associated with cognitive impairment (70). A register study of 1 million term children with low Apgar suggested that five minute Apgar might be as close as we get to 'asphyxia' in register studies as Apgar at 1 minute often is caused by a temporary depression and 10-minute Apgar might not be reflecting intrapartum events and is often missing if five-minute Apgar is 10 (30).

#### 4.2.5.1.2 Umbilical cord pH, acid-base and lactate

Umbilical cord pH and acid-base assessment are more objective outcome measures than Apgar score, and reflect the fetal metabolic condition at birth. In Sweden in 2006-2012 the reporting frequency of pH-data varied from 18.2-97.0%, making it unreliable as a national outcome measure. For infants that later in life show signs of cerebral palsy, data on arterial umbilical cord pH <7.00 or acid-base assessment (i.e. base deficit  $\geq 12$  mmol/l) at birth are one of three essential criteria to establish the association with asphyxia at birth (71). A large cohort study of 51,519 newborns concluded that the threshold for adverse neonatal outcome is pH 7.10 but the risk rises sharply below an arterial pH of 7.00. At pH <7.05 the relative risk of an Apgar score of <7 at 5 minutes was 11.74%. However, a pH below 7.00 accounted for only 20-24% of adverse neurological outcomes, indicating that most neonates with neurological morbidity have normal cord pH values (72). Variables other than asphyxia measured by pH and Apgar are probably responsible for the main part of neurological morbidity of the neonate. These variables, such as chorioamnionitis, and maternal morbidity such as severe preeclampsia, and bleeding, are important to consider when assessing risks in a vaginal delivery (73).

During labour, fetal scalp blood sampling to analyse lactate is used to identify neonates developing metabolic acidosis. Lactate can be analysed in smaller blood volumes than pH and can discriminate metabolic acidosis from harmless respiratory acidosis (74). Although the correlation between lactate, umbilical cord pH and base deficit is considered good (75), lactate is still not regularly used as an outcome measure.

#### 4.2.5.2 AOI and CUSUM charts

Mann et al recognised that the frequency of adverse outcomes in obstetrics is low and thus combined obstetric outcome measures into an Adverse Outcome Index (AOI) where adverse events (i.e. maternal death, admission to NICU, Apgar <7 at five minutes) are weighted to represent the severity of the outcome (59).

Draycott et al claim that outcome measures are the most important to focus on, as process measures can mask poor clinical care (55). Sibanda et al defined a set of clinical indicators covering labour and delivery outcomes using a Delphi process (57). They argue that it is important not only to collect appropriate data but to transform these data into meaningful information for front-line staff using maternity dashboards or equivalents. Continuous monitoring of intrapartum outcomes in CUSUM charts makes early detection of adverse trends possible and if followed by effective interventions, care can improve (56, 76).

### **4.3 THE NATIONAL PERINATAL PATIENT SAFETY PROGRAMME**

#### **4.3.1 Background**

In August 2007 the Swedish National Patient Insurance Company LÖF (LÖF) initiated a meeting with representatives from the three professional associations for obstetricians (Swedish Society for Obstetrics and Gynaecology, SFOG), neonatologists, (Swedish Neonatal Society), and midwives (the Swedish Association of Midwives) to discuss possible interventions against severe delivery-related asphyxia. The initiative was based on preliminary data on recurrent errors in perinatal care, explored in a study of claims due to severe delivery related asphyxia (37). Two projects were designed: 1) the national Perinatal Patient Safety Programme (PPS programme) with the aim to further investigate risk areas and improve current patient safety practices in perinatal care, and: 2) an educational CTG programme.

The overall objective was to improve patient safety in delivery care in order to reduce the number of infants with delivery-related asphyxia.

The programme, engaging all 46 obstetric units in Sweden, started with an introduction meeting at the obstetric unit, followed by local self-assessment, a peer review process, a mutual agreement on improvement measures adjusted for the local setting, a period of local improvement work, and a final report describing achieved improvement measures. The whole process was summarised and discussed at an assembly where good examples were shared among the participants.

A web-based learning programme for education and training in fetal assessment and cardiotocography (CTG) was developed, as a need to improve knowledge in fetal assessment was highlighted (23) (Neoventa Medical®, Gothenburg, Sweden). The programme includes a theoretical background on fetal physiology, methods for fetal surveillance, classification and assessment of CTG and interactive trainings sessions ([www.ctgutbildning.se](http://www.ctgutbildning.se)). Finally there is an option to take a written examination, using a special log-in provided to management functions at the unit. It was launched, free of charge, to all delivery clinics in Sweden in 2009, when local improvement processes were ongoing.

#### **4.3.2 Empirical setting**

Swedish healthcare is tax funded, the Swedish government being responsible for medical care policy and overall health while the responsibilities for executing primary care and specialist care is decentralised to local and regional authorities consisting of 21 county councils. There were 46 obstetric units in Sweden ranging in size from 298 to 7,297 deliveries per year in 2012. When divided into subgroups, 10 units had >3,000 deliveries per year, 21 units had 1,500-3,000 and 15 units had <1,500 deliveries per year,

respectively. The facilities ranged from small local hospitals to large university centres. The number of live births in Sweden during the study period was 764,684, with a 7.24% increase from 104,293 in 2006 to 111,843 in 2012.

#### **4.3.3 Swedish National Patient Insurance Company LÖF**

All patients receiving care within the county councils are insured by the Swedish National Patient Insurance Company LÖF, the largest insurance company for patient injuries in Sweden. Patients can report care related injuries to LÖF who initiates an investigation executed by medical experts. If the injury is considered avoidable, the patient is entitled to financial compensation. Of the 20-50 annual claims for financial compensation due to severe delivery related asphyxia approximately 45% are found to be avoidable (= settled claims). Delivery related asphyxia accounts for more than 20% of LÖF's total indemnifications. The financial compensation to one child with delivery related asphyxia amounts to 8 MSEK (848,000 € as at 8<sup>th</sup> of January 2015), sometimes more. Preventing injuries, based on knowledge that can be extracted from the claims, is one of the assignments of LÖF.

During the Perinatal Patient Safety programme, LÖF provided administrative support, i.e. were responsible for the arrangement of start-up seminars, meetings and hearings, travel costs and equipment needed. A programme administrator co-ordinated all contacts with the managers of obstetric units, the peer reviewers and the steering committee. The programme administrator also collected agreements from local obstetric unit managers to participate under the stipulated conditions and assembled all documents and reports that evolved during the programme process.

The steering committee comprised the presidents and board members of the professional organisations and the Chief Medical Officer (CMO) of LÖF. Meetings were held four times a year, from 2007 until 2011.

#### **4.3.4 Specific interventions used in the PPS programme**

The work process, illustrated in figure 4, was the same for all obstetric units. An experienced midwife, obstetrician or neonatologist and the project administrator (PA) met with the management of the unit, sharing information about the aim and purpose of the programme, as well as methods, time frame and support functions available. The unit management assigned a multi-professional group with representatives from all front-line staff involved in delivery care, including assisting nurses, midwives, obstetricians, anaesthesiologists and neonatologists. They performed a self-assessment using a formulary with selected questions and were asked to share information with enough detail so that issues related to safety were assessable for the external peer review group. They

were also asked to attach relevant guidelines and routines, and present ideas for improvements emerging during the process (Figure 4).

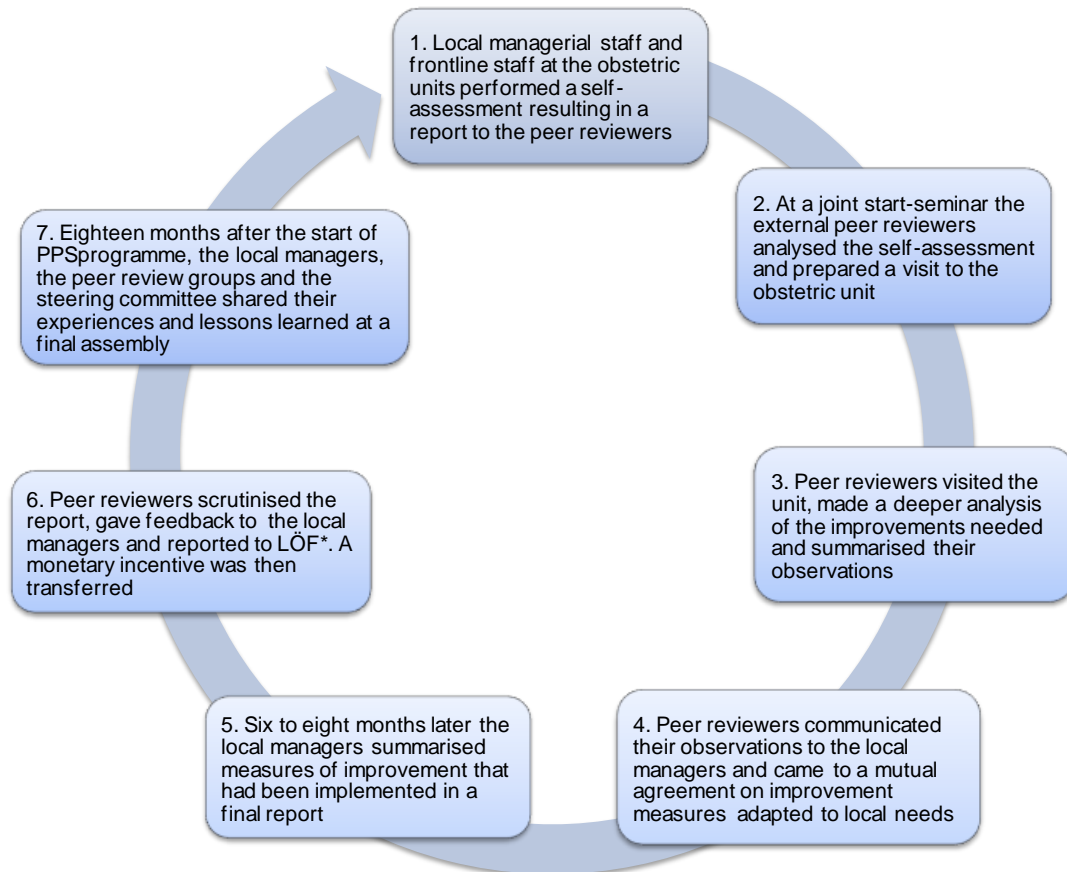


Figure 4. The work process of the PPS programme

The peer review group took a part of the results of the self-assessment and scrutinised it before a visit to the clinic, all within four months from the introductory meeting. After the visit they summarised their observations in a written feedback to the clinic. Then the unit management and the peer-reviewers came to a mutual agreement about a number of improvement measures for the unit to focus on, in order to meet specific needs of improving patient safety for the infant during delivery.

Six to eight months later the obstetric unit management wrote a final report to the peer reviewers and the programme administrator on what measures of improvement had actually been implemented.

#### 4.3.5 The self-assessment instrument

Self-assessment is used to explore base line performance and compare it to reference standards. It is a systematic review that can allow staff to better understand the organisation's results and activities, illuminate strengths and weaknesses and point out areas for improvement (71). Self-assessments can reveal many possible areas of improvement and be of help when prioritising among these, support change processes and provide structured support to ensure sustainability in new practices for successful change. Regular self-assessments can ensure that well-adapted methods are developed (72).

The result of a self-assessment is often a report that can be used as basic data for an external reviewer participating in a collaborating process with the organisation. This mixed model process can provide helpful feedback and further stimulate continuous improvement efforts.

Before the start of the PPS programme, the self-assessment protocol was constructed by an expert group, including an obstetrician, two midwives and a neonatologist, all recruited by the professional organisations. The experts investigated the gaps between best and current practice, drawing from their clinical experience. By critically evaluating every step of the pregnant woman on her way through the labour and delivery process, from the first telephone call to the obstetric unit until after delivery and with focus on safety for the infant, the possible risk areas were listed as follows: organisation, communication, competence, drug administration, medical technique, documentation and follow-up (Figure 5).

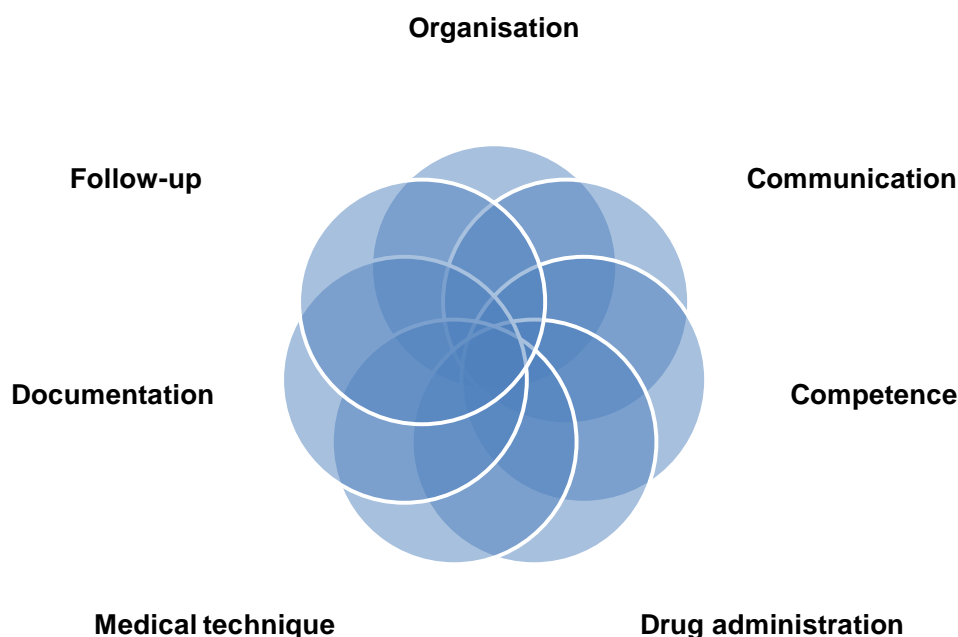


Figure 5. Risk areas in obstetric care.

Based on these identified risk areas, questions to stimulate self-reflection in the local setting of the obstetric units were constructed. The essences of the questions were:

“What routines or guidelines regarding area x are in place in your obstetric unit?”

“How do you make sure that these routines are followed?”

All risk areas were approached by both questions. The questions were phrased in an open-ended manner with the aim of being the starting point for a multi-professional discussion, leading to increased awareness of measures related to the safety of the infant during delivery and to promote insights about contingent brittleness within the system. Involving obstetric staff from all levels of care ensured that management as well as front-line perspectives were considered and that consciousness of the features of the programme was disseminated throughout the organisation. The questions were also phrased to help the units describe their local context and conditions with such clarity that the peer review group could get a picture of potential areas of improvement. The final self-assessment tool contained a total of 23 questions (Figure 6). The units spent approximately eight weeks working with their self-assessments and the resulting reports were then sent to the peer review group to be further analysed prior to their onsite visit.

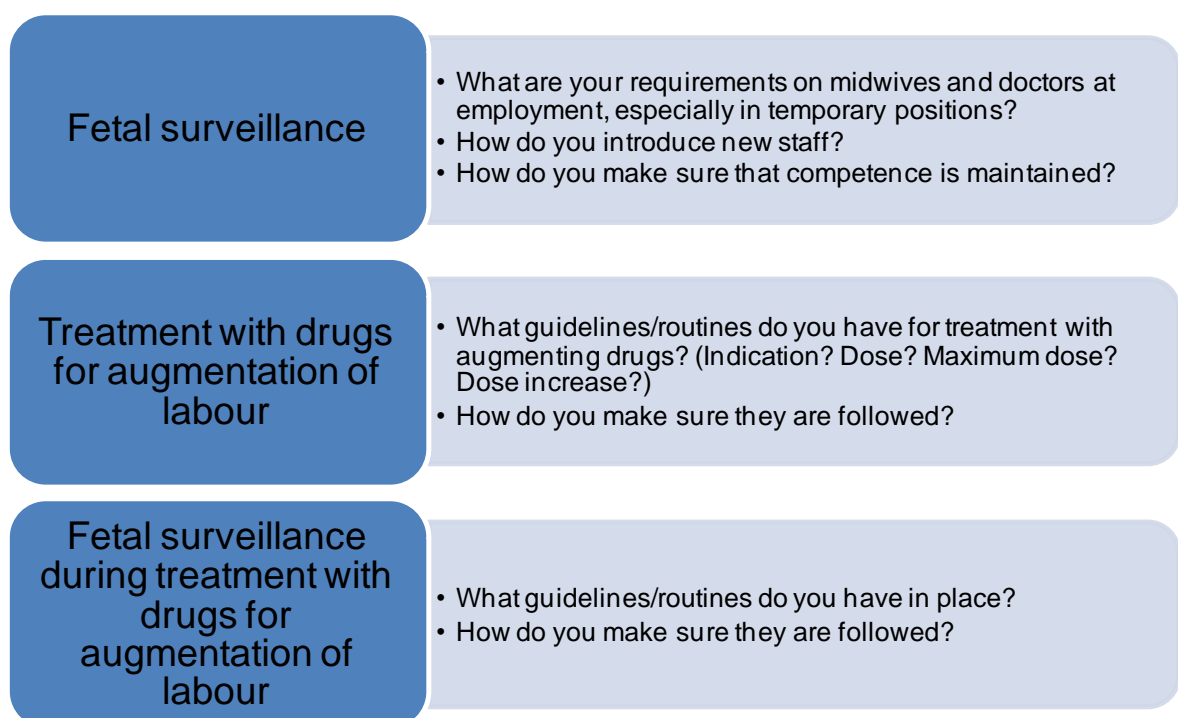


Figure 6. Examples of questions considering two of the risk areas, as phrased in the self-assessment tool.

The professional organisations were responsible for the content of the self-assessment tool and had full liability for its development.

#### **4.3.6 The peer reviewers**

Peer review is the systematic assessment of the performance of an organisation or department by a group of professionals, with the objective to help the target organisation to improve, implement evidence-based practice, and adhere to current standards (77). When peer review is performed by professionals working in the same field as the organisation that is being reviewed, there is a good chance that learning is mutual. However, peer review alone is not likely to be effective in supporting improvement processes, and internal quality monitoring is an important foundation for achieving synergetic effects (78).

Eighty-one peer reviewers (29 obstetricians, 23 midwives and 29 neonatologists) were recruited by the professional organisations. The peer reviewers participated from one up to four revisions each. From the second sequence and onwards the peer review process was designed so that reviewers with previous experience worked with new reviewers. The teams consisted of an obstetrician, a midwife, and a neonatologist and they co-operated with one obstetric unit per sequence. Between 13 and 15 peer review teams were gathered at a start-up seminar. They were introduced to the project and their assignment, then they scrutinised the obstetric units' self-assessments with the help of a structured protocol. They were asked to check if the questions were comprehensively answered and to phrase new questions about details in need of clarification. The instruction was to support the obstetric unit management in their efforts to improve patient safety within existing resources, to have an objective approach and yet to be explicit about observed risk areas. Each peer review team planned their visit to the obstetric unit and at the end of the start-up seminar all peer review teams assembled to discuss general issues of concern about the peer review process and important patient safety issues.

#### **4.3.7 Visits to the departments, feedback reports and mutual agreement on improvement measures**

The external peer review team visited the obstetric unit about a month after the analysis of the self-assessment, with the visit lasting one or two days. The peer reviewers met with representatives from all frontline staff and were guided through facilities to get a deeper understanding of the local setting. At the end of the day they met with the management, sharing observations and thoughts. Within four weeks after the visit the peer reviewers sent a protocol-based feedback report to the unit management. The feed-back contained comments on strengths, and on areas of possible improvement in general and in detail with regard to the 'risk area questions' in the self-assessment tool. Finally, at the end of the feed-back report the peer reviewers summarised the most important areas to improve.

The unit managers were invited to respond to the comments and this communication between the peer review team and the management eventually led to a mutual agreement on a number of improvement measures to realize.



Six to eight months later the units sent their final report to the peer-reviewers and the programme administrator, describing details on improvement measures that had been accomplished.

#### **4.3.8 Time frame of the Perinatal Patient Safety programme**

In spring 2008 a pilot intervention process was carried out in four obstetric units of different sizes and with different prerequisites. The self-assessment tool and a peer review process were tested. The self-assessment tool then consisted of 23 questions. The pilot process was received positively and after some minor clarifications in the instructions to the peer reviewers and improvement of the self-assessment tool resulting in a total of 27 questions, all delivery clinics in Sweden were invited to participate.

After the pilot sequence engaging four units, the first sequence started in September 2008 and included 14 delivery clinics, followed by sequence II in January 2009 with 13 obstetric units, and sequence III in September 2009 with 15 obstetric units (Figure 7).

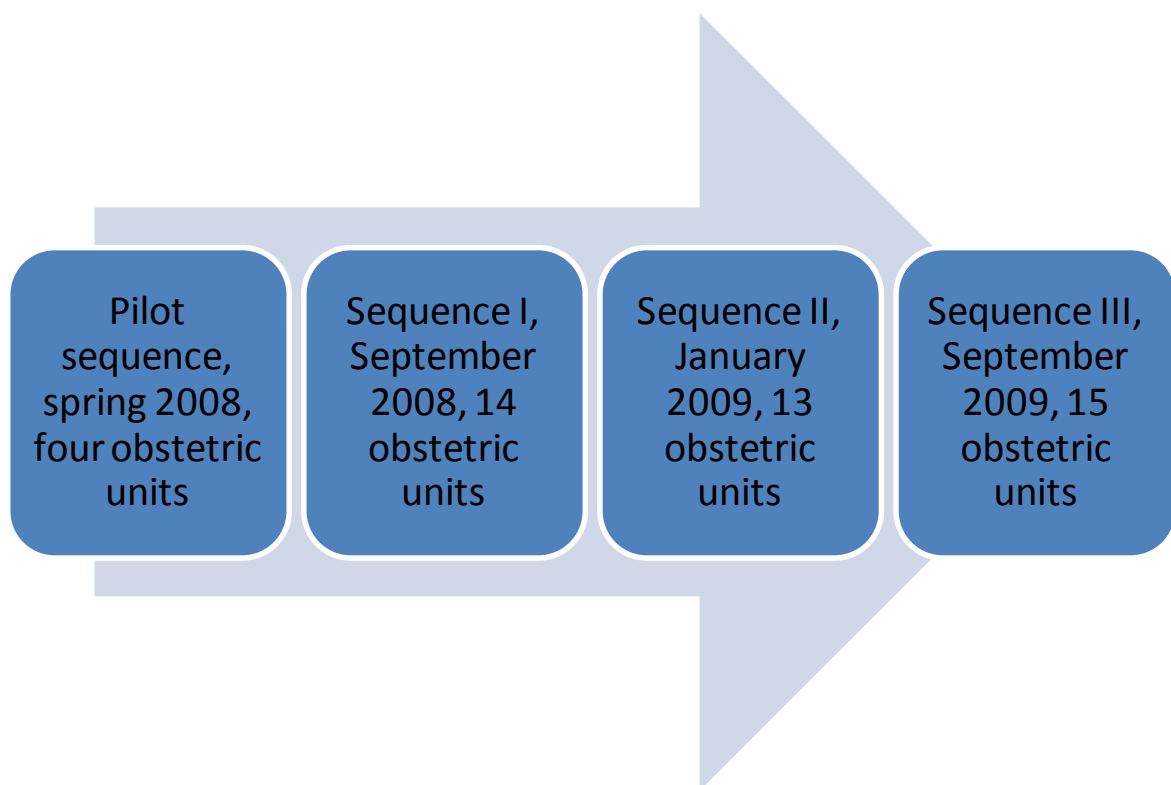


Figure 7. Starting points for each sequence in the PPS programme

## 5 AIMS OF THE THESIS

The overall aim of this thesis is to explore the implementation and results of a national patient safety intervention aiming at improving outcomes for the newborn infant.

The specific aims of the studies were:

- To investigate how the initial programme interventions, including self-assessment, peer review and agreement for change, affected the teams and their mental models of patient safety improvement (Paper I).
- To evaluate if a web-based learning programme could bring about a higher degree of individuals who correctly classified cardiotocography (CTG) recordings in a non-selected population of midwives and physicians (Paper II).
- To study if pair-wise interpretation of CTG could bring about a higher level of correctly classified CTG-recordings compared to individual interpretations in a non-selected population of midwives and physicians (Paper III).
- To assess the results of the PPS programme by addressing local improvement measures taken, changes detected in the proportion of term newborns with asphyxia and changes in the frequency of asphyxiated newborns, as measured by LÖF-settled claims (Paper IV).
- To investigate the risk of being subjected to factors associated with suboptimal care during labour and delivery by performing a criterion-based review (Paper V).

Figure 8 provides an overview over the papers in relation to the PPS and CTG programmes, represented in an input-process-output model. Paper I evaluated the input of the core interventions on a national level, Paper II and III evaluated output data as in educational effects on a local level, Paper IV evaluated national outcome data (Apgar and settled claims) and local improvement efforts reported in the final reports and Paper V evaluated local output data before and after the PPS programme.

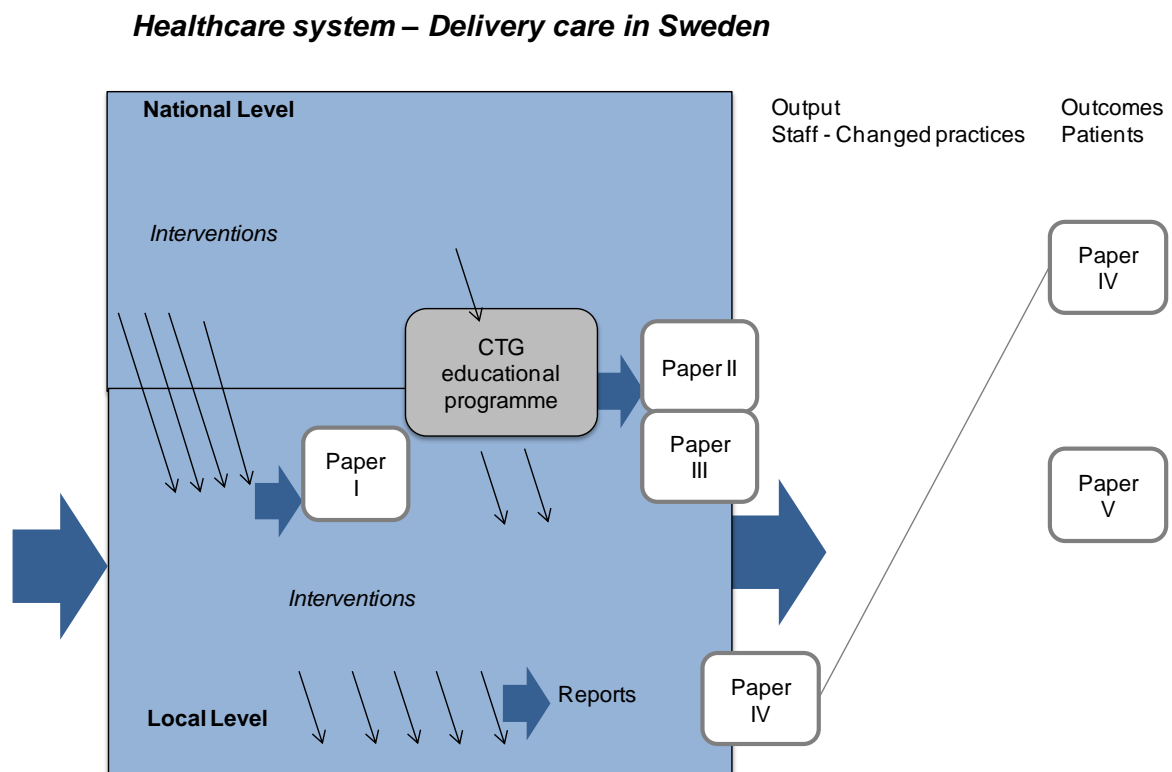


Figure 8. An overview of the five papers in relation to the PPS and CTG programmes

## 6 MATERIAL AND METHODS

Table 1. Overview of the general design of the project.

Paper	Design	Data collection	Participants	Analysis
I	Qualitative interview study of multiprofessional teams	Semi-structured telephone interviews	80 management directors, head midwife and senior consultant in obstetrics participating in the PPS programme	Content analysis using a priori coding (deductive content analysis)
II	Before and after-study, inclusion from Sept. 2009 to Apr 2010.	Assessments of CTG-recordings before and after a web-based education programme.	179 midwives and doctors at the obstetric unit, Södersjukhuset, Stockholm, before and 135 after education.	Fishers' exact test
III	Comparative study between individuals and pairs classifying ctg, inclusions from May 2012 to Sept. 2012	Assessment of CTG-recordings individually and pairwise, respectively.	387 midwives and 149 physicians at six obstetric units in Stockholm	Wald's test to compare proportions
IV	Multi-method retrospective evaluation, final reports 2008-2011, Apgar data from all Swedish units 2006-2012 and claims settled at LÖF 2000-2014 and q1 2015	a) Final reports on improvement measures from the units b) Apgar data from patients with AS<7 <sup>5</sup> c) Settled claims from LÖF	a) All delivery dept. in Sweden b) All 414 978 term newborns born in Sweden 2006-2012 c) The number of settled claims granting economic compensation by LÖF due to severe delivery-related asphyxia	a) Content analysis b) Mantel Haenszel technique c) Segmented logistic regression analysis and content analysis
V	Comparison of two case-control-studies with respect to a criterion-based review of care processes during delivery in 2004-06 and 2009-12	Data on maternity and delivery care from obstetric records in Obstetrix® (Siemens), CTG tracings recorded during labour and neonatal records were collected according to a protocol	336 cases with AS <7 <sup>5</sup> , born in Stockholm, 336 controls with AS 10 <sup>5</sup> . Cases compared with cases, controls compared with controls	Fishers' exact test and multivariable logistic regression analysis, first unadjusted and then adjusted for all background variables

## 6.1 PAPER I

### 6.1.1 Study population and design

Study I is a hospital-based qualitative study based on 80 semi-structured telephone interviews with the clinical department head of the obstetric units and the multi-professional teams i.e. head midwives and senior consultants, in 27 obstetric units taking part in the first and second sequence of the programme. The questions in the interviews sought to elucidate how the process of self-assessment, the peer review (including the visit to the obstetric unit), the written feedback report, and the agreement for change were perceived and if this intervention process had any impact on team mental models and readiness for change at the obstetric units.

## 6.2 PAPERS II AND III

### 6.2.1 Creation of a 'reference standard' CTG pool

A pool of 'reference standard' CTG-recordings was created with the help of four experts (two senior midwives and two senior consultants in obstetrics) (Figure 9).

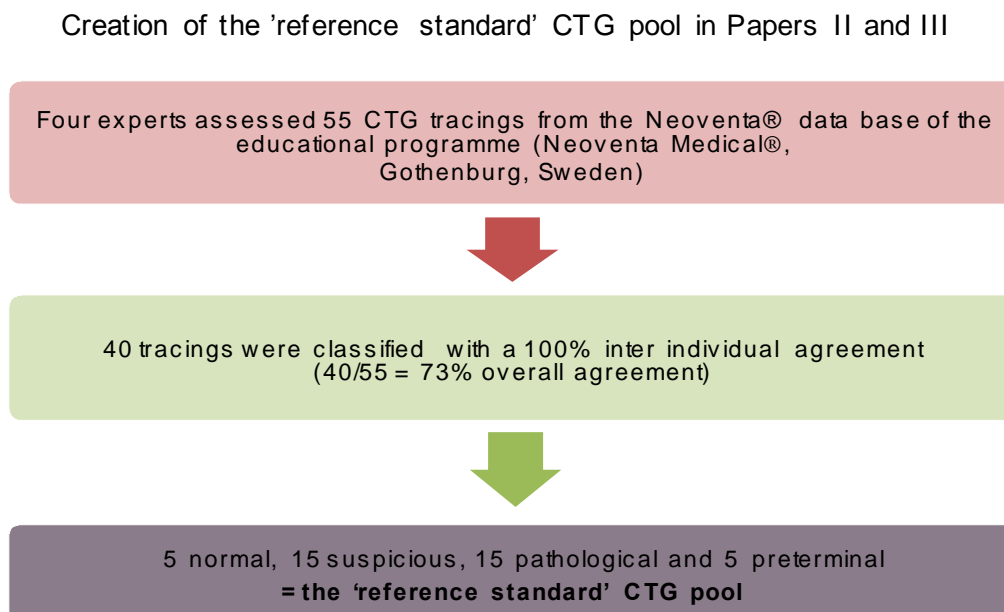


Figure 9. The procedure for creating a 'reference standard' CTG pool

## 6.2.2 Paper II

### 6.2.2.1 Study population and design

The setting was the largest on-site obstetric unit in Sweden (Södersjukhuset, Stockholm), delivering approximately 7,500 infants a year, and providing perinatal care to both high and low risk pregnancies and deliveries. All midwives and physicians taking active part in front-line work were invited to participate. Of the eligible 180 individuals, 179 classified one randomly chosen CTG-recording before getting access to the web-based programme. Four hours of preparation time with full salary was provided for individual studies in the programme before performing the examination. Seventy-five percent (n=135) completed the education and passed the examination and all 135 participated in the second assessment of a CTG from the pool. A third assessment was performed during the same timeframe as the second, allowing 55 couples (midwife-midwife or midwife-physician) to assess a CTG-tracing together. Forty-four individuals claimed that the time allotted for studies in the programme was too short and declined further participation.

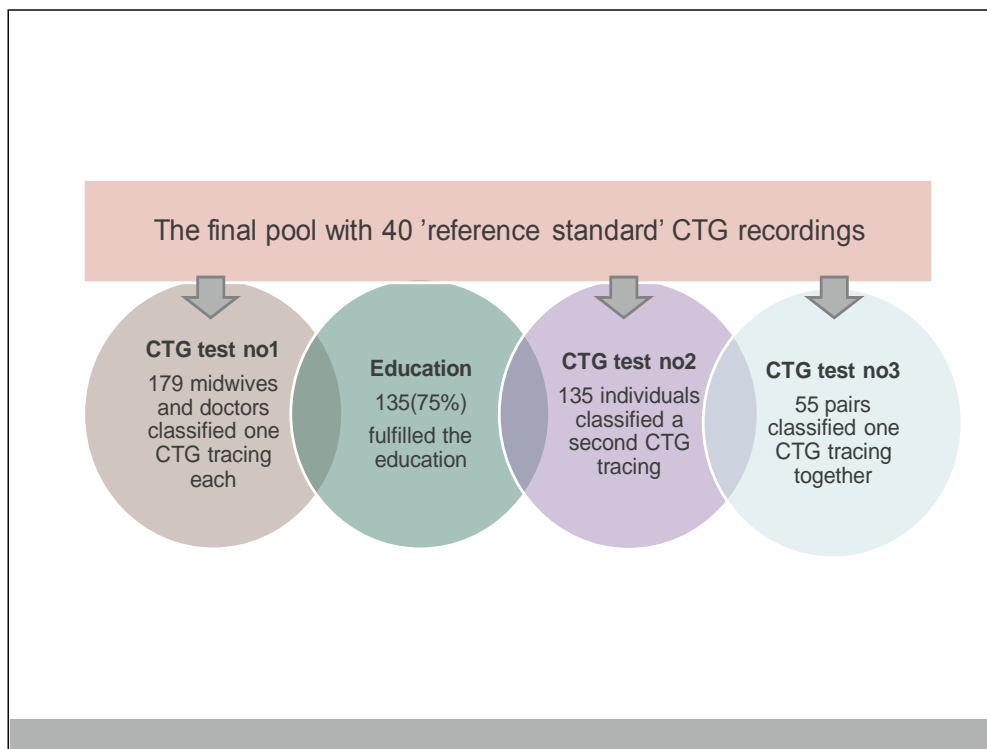
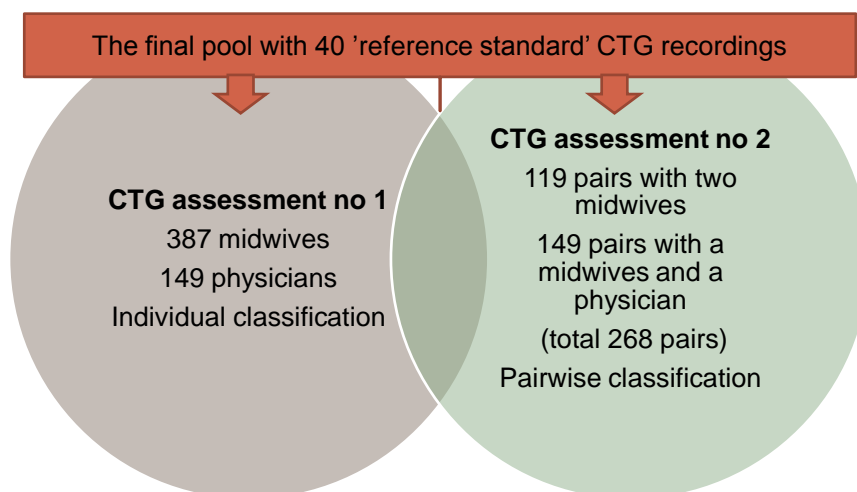


Figure 10. Study process, Paper II.

## 6.2.3 Paper III

### 6.2.3.1 Study population and design

The setting was five obstetric units in the Stockholm region and Uppsala Akademiska Sjukhus, Sweden, each delivering between 1,500 and 7,500 infants a year. Five hundred and thirty six midwives and physicians were recruited. First, each participant individually classified a randomly selected CTG-recording from the pool of 'reference standards', followed by the classification of a second randomly selected CTG-recording made as a pair with a colleague, generating 149 pairs of a midwife and a physician and 119 pairs of two midwives (Figure 11).



3

Figure 11. Study process in Paper III.

## 6.3 PAPER IV

### 6.3.1 Study population and design

In this multiple methods, observational retrospective evaluation on final reports 2008-2011, Apgar scores 2006-2012 and settled claims 2000-2014, the outcomes were a combination of qualitative and quantitative data:

- Final reports describing local improvement measures were retrieved to explore the different improvement strategies at the departments.
- Asphyxia defined by low Apgar score at birth. Information about 414,978 newborns in 2006-2012 with Apgar score <7 at five minutes of age was collected using the national Medical Birth Registry (MBR) (n= 3622). We included single, term ( $\geq 37+0$ ) fetus, alive at mother's admission for delivery with spontaneous or induced start of labour. The units were divided into three groups according to incidence of Apgar score <7. Apgar score was extracted for three periods for each unit: before (I), during participation (II) and after PPS programme (III) (figure 12). Initially pH-data was collected to be analysed together with Apgar scores. The inconsistency in reporting frequency made it impossible to analyse pH-data with sufficient power. We therefore chose to exclusively analyse Apgar score in all children born in Sweden.
- The frequency of settled claims as a result of severe delivery-related asphyxia granting financial compensation from LÖF in 2000-2015 was analysed exploring possible changes in trends over time.

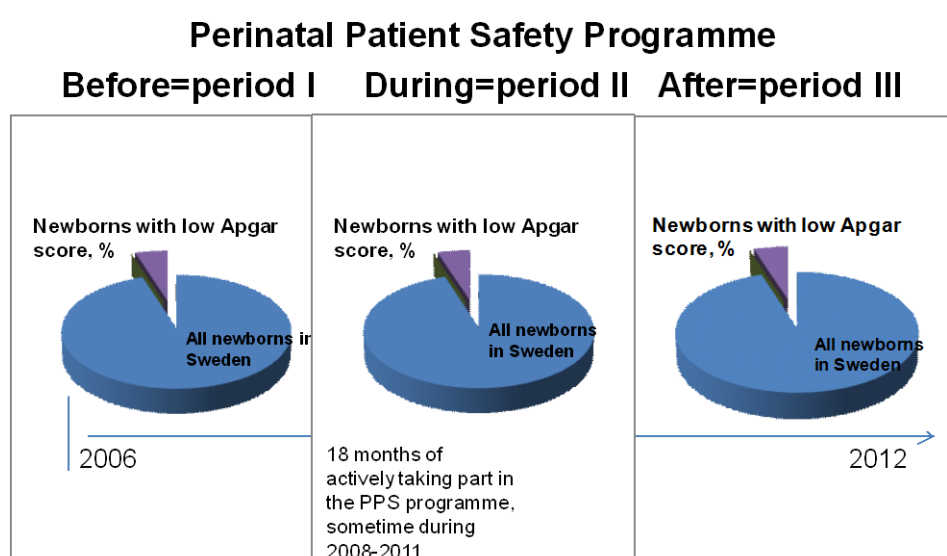


Figure 12. Timescale for Paper IV.



## 6.4 PAPER V

### 6.4.1 Study population and design

This is a case-control, criterion-based review of care processes during labour and delivery in six obstetric units in Stockholm County during two periods (2004-2006 vs. 2009-2012). The set of criterion used were identical in both reviews; however, only the criterion considered having low inter-rater variability were used when comparing the two study periods. The incidence of Apgar at five minutes from the two periods was retrieved from the MBR.

Infants born in Stockholm 2009-2012 were included. Cases were infants born with Apgar score <7 at five minutes of age at gestational age of  $\geq 33+0$ . For each of the 336 cases, one randomly selected control with Apgar 10 at five minutes of age was retrieved and matched for hospital and year of birth. The matching for hospital was made as we had no intention to assess differences between hospitals in Stockholm.

The data used as comparison was a population-based case-control study of infants born in Stockholm 2004-2006 (37). Inclusion- and exclusion criterion were identical. Cases born after the PPS programme (i.e. 2009-2012) were compared with cases before the PPS programme (i.e. 2004-2006) and controls were compared with controls (Table 2).

Cases	Controls
Infants with gestational age $\geq 33+0$	Infants with gestational age $\geq 33+0$
Spontaneous or induced onset of delivery	Spontaneous or induced onset of delivery
Normal CTG tracing at admission to the maternity unit	Normal CTG tracing at admission to the maternity unit
Apgar score < 7 at five minutes	Apgar score 10 at five minutes
	Matched for year and hospital

Table 2. Inclusion and exclusion criterion for the population in 2009-2012. Criterion were identical with the population of comparison from 2004-2006.

Cases were identified by an obstetric record system Obstetrix® (Siemens), used by all obstetric units in Stockholm. Controls were identified by the Swedish Medical Birth Registry (MBR). In Stockholm, all antenatal and obstetric information collected into the registers is retrieved from Obstetrix® (Siemens).

We excluded infants with gestational age  $GA \leq 33+0$  as prematurity in itself is a major risk factor for low Apgar, regardless of the handling during delivery (30, 37). Six hundred fourteen cases were identified; 105 infants were excluded due to intrauterine death diagnosed before onset of labour, 112 due to non-reactive CTG at admission; 48 due to planned or acute cesarean section before start of labour, and 13 due to missing records or CTG-tracing, respectively.

A structured protocol, (page 75) (79), was used (except for the data on detriment number and exclusion criterion) . The protocol has previously been used in four studies aiming to describe the most common risk factors and causes of substandard care during labour, (23, 37, 80, 81). Data from obstetric and neonatal records was transcribed into the protocol. Details about maternity and delivery care were abstracted and CTG-tracings recorded during labour were scrutinised by the author. Information was collected about onset of labour, interventions during delivery, engagement of physicians, pH/lactate if assessed, mode and time of delivery, presentation, gestational age, gender, Apgar scores, umbilical cord tests, resuscitation activities, and care and diagnoses in the neonatal unit.

All CTG-tracings were scrutinised by one of the authors, Charlotte Millde Luthander (CML). The inter-observer variability between Sophie Berglund and an external expert (Professor Ingmar Ingmarsson, senior consultant, †) in the previous study was 86 % when one hundred randomly selected intrapartal CTG-tracings were assessed. The procedure was repeated, showing an inter-observer variability between CML and SB of 90%.

#### **6.4.1.1 Criterion used for comparison when assessing quality of care**

Fetal surveillance during labour in accordance with Swedish guidelines was defined as follows: on admission to the obstetric unit CTG is performed for at least 20 minutes. If CTG is normal and the pregnancy is uncomplicated, intermittent CTG is performed every two hours during the first stage of labour and the midwife registers fetal heart sounds every 15-30 minutes before, during and after contractions. During the first 45 minutes of the second stage of labour, fetal heart sounds are registered every fifth minute (74). Continuous CTG is recommended if active pushing exceeds 30 minutes due to an increased risk of asphyxia (82, 83). In cases of suspicious or pathological CTG, continuous CTG is required, as well as if maternal risk factors are known before delivery, such as diabetes, severe pre-eclampsia,  $GA \leq 36+6$  or  $\geq 42+0$ , suspicion of placental insufficiency or intrauterine growth restriction. It is also a requisite if complications occur during delivery, such as meconium stained amniotic fluid, bleeding, or maternal fever  $>38.0^{\circ}\text{C}$ . If treatment with oxytocin due to inertia (defined as no progress of labour for three hours after an initial normal progress) is initiated, simultaneous surveillance of contractions is indicated to detect uterine hyper stimulation (i.e.  $\geq$  six contractions per 10 minutes) (83). According to our national guidelines, fetal blood scalp sampling (FBS) should complement fetal surveillance if the

CTG-recordings are intermediate or pathological. In accordance with the previous study a time limit of 40-60 minutes was used to define when FBS was indicated. If lactate is  $<4.8$  mmol/L or the pH is  $<7.20$  but the CTG is continuously intermediary or pathological, a follow-up of FBS is recommended every 20-30 minutes depending on progress and stage of delivery (37).

When scrutinising the medical records, detailed information such as at what time the CTG became pathological, if and when fetal blood samples were taken, and the time of delivery, was collected. However, the overall judgment of 'substandard care' used in the previous study (37) was a synthesis of all measures and actions taken during delivery. For example, if FBS was used in accordance to guidelines to assess fetal lactemia, or delivery was expedited, deliveries with pathological CTG for more than 45 minutes before delivery were considered correctly handled. We considered that this composite assessment increased the risk for high inter-rater variability and chose to include only the objective measures with dichotomous answers. Thus we defined three categories of increased risk for adverse outcome during labour and delivery and a fourth category indicating if any of the included risk factors were present:

- 1) Neglecting to supervise fetal well-being was defined by no CTG-recording for more than 2.5 hours after admission test and/or more than two hours between CTG-recordings, and/or non-interpretable CTG due to poor quality, and/or no fetal blood sample (FBS) despite a continuous intermediary or pathological tracing for more than 45 minutes, and/or no follow-up of FBS despite continuous pathological CTG with a normal FBS, or no follow up of pre-acidotic FBS (according to stage of delivery).
- 2) Incautious use of oxytocin was defined by uterine tachysystole as six contractions or more every 10 minutes for more than 20 minutes, and/or increased dose of oxytocin despite pathological CTG or tachysystole.
- 3) A complex vaginal instrumental delivery was defined as an inappropriate trial of labour with a vacuum extractor or forceps in the following circumstances: incomplete cervical dilation, non-cephalic presentation or cephalic malpresentation, non-engaged fetal head, a clear indication of cephalic disproportion, extraction exceeding 20 minutes or more than two cup detachments.
- 4) The risk of any factor indicating increased risk was a summary of one or more of the above defined risks.

## **6.5 STATISTICAL ANALYSES**

### **6.5.1 Paper I**

The qualitative data resulting from Paper I was analysed using directed content analysis (84). This deductive analysis used a priori coding in which certain patterns of text segments were thought to correlate to the development of team mental models (TMM) and a readiness for change. Each text segment was coded according to which of the three interventions it related to (self-assessment, peer review, and agreement for change), in what way it was modelling TMM, and if it had a positive or negative value. The respondents' results were compared with the ideal case, i.e. codes indicating that development of TMMs had taken place or actions were taken to improve patient safety.

### **6.5.2 Paper II**

Fisher's exact test was used to compare the proportion of individuals classifying CTG correctly before versus after the education, (significant if  $p < 0.05\%$ , two-tailed).

### **6.5.3 Paper III**

Based on the results in Paper II, we assumed that the level of correctly classified CTG would be 65%. Wald's test was used to adapt to the fact that analyses are based on clustered data where the same individual appears twice in the data, once individually and once in a pair. To detect a 15 percentage point (80% correctly classified CTGs among pairs), the sample size needed was 151 individual observations and 151 pairwise observations. Wald's test was also used for subgroup analysis between units and between pairs consisting of midwife/midwife and midwife/physician.

### **6.5.4 Paper IV**

In Paper IV we used three different statistical analyses.

a) A content analysis in which the text in the final reports was analysed in several steps, inductively coding text segments into 14 categories of 'type activities'. The number of units reporting each category was noted and the percentage of all 46 obstetric units was calculated.

b) Adjusted odds ratios for Apgar score  $<7$ , Apgar score 0 – 3, and Apgar score 4 – 6, for period III vs. period I, were obtained using the Mantel Haenszel technique. Stratification was made for hospital, maternal age, parity, smoking, and BMI.

c) Data on the births of one-quarter for newborns granted compensation in 2000-2015 were analysed using segmented logistic regression. The purpose was to explore possible trend changes in relation to the PPS programme. The analysis was exploratory and hypothesis generating (85).

#### **6.5.5 Paper V**

We used Fisher's exact test to study the difference between cases before and after the PPS programme, and controls compared with controls for each of the background factors (Table 1). We used multivariable logistic regression to study the association between each of the four categories of risk factors between the time periods, first unadjusted and then adjusted for all background variables. We used the Hosmer-Lemeshow test to evaluate the validity of logistics.

## **6.6 ETHICAL APPROVALS AND CONSIDERATIONS**

Ethical approval was obtained from the Regional Ethics Committee in Stockholm (D: nr 2010 -1603 31 -4). It considered the evaluation of the PPS programme as activity development and declared in its advisory statement that there were no ethical objections to the project.

For study IV, an additional complimentary application for approval to analyse data from LÖF was obtained (D: nr 2015 -0152 -32)

For study V a separate application to the Regional Ethics Committee in Stockholm was created and granted (D: nr 2011/573-31/1 concerning cases and controls 2009-2011) as this study involved scrutinising medical records. However, the inclusion of cases and controls from 27<sup>th</sup> May 2009 to 31 December 2011 did not result in large enough samples. Therefore, a complimentary application was made to obtain approval to continue including cases and controls until 31 December 2013 (D: nr 2012/2241-32).

Informed consent was not obtained from individuals participating in the study. There was a risk that the studied individuals might feel that their integrity had been violated, participating in the study without expressing consent. It also is possible that information about the study might have been perceived as offensive for some, concerned as to why nothing had been done before. We can surmise that among the studied cases there is probably a small group of severely injured children, sometimes due to suboptimal care; the suffering for these families is always extensive, in regard to humanitarian values and sometimes economically.

Moreover, there was a risk of uneven participation in cases and controls. The overarching aim of the study was to improve delivery care through systematic identification of factors related to suboptimal care. This required a high participation rate in both groups and a selective drop-out, especially from the most severely injured, would have jeopardised the validity of the study.

It can be argued that to not systematically learn from available data to improve care would be unethical. Therefore we chose to compile de-identified data without informed consent from the individuals participating in the study; this was approved by the Ethics Committee.

## 7 RESULTS

### 7.1 MAIN FINDINGS - PAPER I

*The self-assessment* process was perceived to have illuminated areas for improvements and pointed out strengths that the team or the obstetric unit possessed. New structures of knowledge and a mutual reference i.e. a team mental model emerged during the process and became more or less internalised by the participants.

*The peer review* process was appreciated and welcomed by the majority of participants. The visits by the peer panel were viewed as “enjoyable and exciting” without the feeling of being investigated. It was described as an exchange of knowledge and experiences, gave encouraging feedback, and confirmed the teams’ ideas. The co-operation with the peer review team confirmed and strengthened the teams own ideas and conceptions, rather than adding new ones. Some negative aspects concerning suggestions about unfeasible measures were sometimes perceived as incompatible with the pre-requisites of the unit.

*The written feedback report* was also appreciated as it strengthened and confirmed established conceptions. In some cases the report was delayed, causing frustration and loss of focus.

*The agreement of change* provided motivation by putting some pressure on the units to accomplish the agreed improvements. Some units experienced that the peer reviewers had unrealistic expectations on what was possible to accomplish within the existing healthcare organisation.

### 7.2 MAIN FINDINGS - PAPER II

The experts agreed on 73% of the 55 CTG tracings. Stratification showed that normal tracings were assessed with 100% agreement in 83% of cases, suspicious 76%, pathological 77% and preterminal 44%.

Study participants classified the tracings in accordance with the experts in 64% before the training and 66% after ( $p=0.756$ ). This baseline level of knowledge was higher than expected and close to the level of the experts. Normal tracings were correctly classified in only 36% before, but raised to 80% after training. Pathological and preterminal tracings were correctly classified to a high degree both before and after training (83% vs. 85%, and 70% vs. 67%). All residents included passed the examination within the stipulated six months, and so did 75% of the consultants and 77% of the midwives on the delivery unit. Midwives in the outpatient unit and antenatal ward participated in the training and examination in various degrees, from 17- 60%.

The best result was obtained when CTG was scrutinised pairwise; 79% assessed CTG correctly. The pairwise assessment was only made after the education and was not

intended to evaluate educational effects. However, the result generated the hypothesis that co-operation when scrutinising CTG might improve the quality of the assessment; this was further investigated in Paper III.

### **7.3 MAIN FINDINGS - PAPER III**

The proportion of correctly classified CTG-tracing when scrutinised individually was 75%; when participants cooperated in pairs the proportion was 80% ( $p=0.12$ ). No significant differences were seen between couples consisting of midwife/midwife or midwife/physician. There were large differences between units; maternity unit C showed the highest improvement (74% versus 90%;  $p < 0.01$ ) and maternity unit E had the highest individual baseline (93% versus 77%;  $p = 0.06$ ), table 1, Paper III. The larger but non-significant proportion of correctly classified CTGs among the pairwise classifications persisted after stratification by CTG type.

### **7.4 MAIN FINDINGS - PAPER IV**

All known risk areas were addressed from a multitude of angles; a broad spectrum of measures of improvements to make delivery safer for the infants was described. The incidence of Apgar score  $<7$  at five minutes on a national level remained constant during the study periods. The 15 units with the highest rate of Apgar score  $<7$  at five minutes showed a significant decrease in Apgar score 4-6 at 5 minutes during study period 3, whereas units with the lowest rate of Apgar score  $<7$  showed a significant increase in Apgar score  $<7$  after the intervention. Cases of severely asphyxiated infants due to substandard care showed a declining trend when data on settled claims from LÖF was analysed.

#### **7.4.1 Results of the content analysis**

The final reports varied in structure and length, covering one to 14 pages (mean 3.6, median 4.5) and with 0 to 34 attached documents (mean 6.33, median 3). The reports contained a description of the improvement areas and activities agreed on, followed by more specific accounts of how these measures were realised. Detailed descriptions of changed guidelines and new routines were attached in 62% of the reports. The maternity units agreed on three to 18 improvement measures (mean 7.2). In the final reports, completion of these measures varied from three to 16 (mean 7.9).

Structural measures received most attention, with 83% of units reporting that they had revised, updated and/or implemented new guidelines and routines. The most common subject was guidelines regulating the handling of oxytocin, but routines regulating the management of CTG and alarm situations were often prioritised measures of improvement. Team training was applied in virtually all obstetric departments, but approaches differed from short resuscitation simulations to full scale, multi-professional,



video-recorded communication simulations. Improved documentation routines in situations such as threatening asphyxia at delivery and the introduction of the web-based CTG education programme (n=17, 37%) were other examples of structure measures introduced.

New process and/or outcome measures were reported to have been introduced in 46% of the units. Random sampling to check adherence to guidelines in general, the use of oxytocin, the incidence of umbilical cord sampling, and the classification and documentation of CTG in particular were the focused subjects of attention. Systematic follow-ups of outcome measures such as delivery related asphyxia defined as Apgar <7 at five minutes and regular reviews of quality indicators, were also introduced (Figure 13).

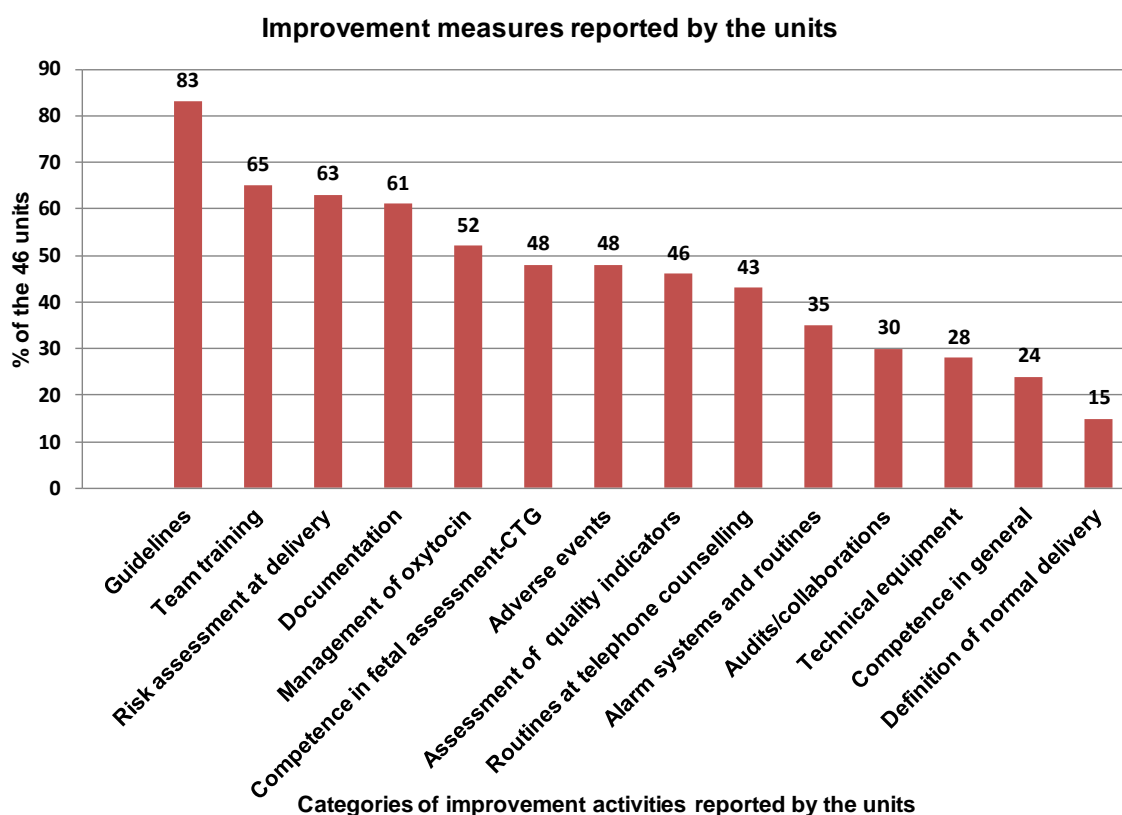


Figure 13. Improvement measures reported by the obstetric units.

#### 7.4.2 Apgar score

The incidence of Apgar score 0-3 at five minutes remained unchanged at 0.2% during the study period. The incidence of Apgar score 4-6 at five minutes was 0.7% during periods I, II and III. The incidence of Apgar score  $\geq 7$  was constant at 99.1% throughout all study periods. The hospitals were divided in three groups according to rate of low Apgar in period I. The odds ratio for an Apgar of 4-6 at the hospitals with the highest frequency of low Apgar was significantly reduced in period III compared to period I, OR 0.84 (95% CI

0.71-0.98). Among the hospitals with the lowest frequency of low Apgar in period I, an increase of low Apgar rates over the study period was observed: OR 1.34 (1.14-1.58).

### **7.4.3 Settled claims**

The quarterly cumulative incidence of settled (i.e. financially compensated) claims showed a slightly decreasing trend (0.5%;  $p=0.279$ ) from 2000 to 2012, the programme's final year, and this measure decreased further in 2012-2014 (7.5%;  $p=0.049$ ), (figure 2).

## **7.5 MAIN FINDINGS - PAPER V**

Changes in background factors among cases were an increased number of primiparas, more women with a previous cesarean section, and an increased incidence of infertility and EDAs in the later period. Among infants born with Apgar 10 at five minutes, there were less instrumental deliveries and a change in the pattern of birth weight, with more infants in the lower birth weight ranges. The incidence of Apgar <7 at five minutes was 8.0 per 1000 (95% CI 7.3-8.7) before, and 8.2 (95% CI 7.6-8.8) after the PPS programme.

In the criterion-based review of infants with Apgar 10 at five minutes (controls), guidelines regulating fetal surveillance during labour were neglected more often in the later period [aOR 2.3 (95% CI 1.56-3.43)]. Factors contributing to this increase were that intervals longer than 2.5 hours between CTG-recordings were more common and that CTG-recordings more often were non-interpretable. Fetal blood samples were indicated but not performed in 8.6% in 2004 -2006 and in 8.8% of the deliveries in 2009-2012 (ns).

Increased doses of oxytocin despite pathological CTG also became more common; however, data about increased doses of oxytocin was more often missing in the later period (p-value <0.01), (table 3, Paper V).

Among infants with Apgar <7 at 5 minutes (cases), there was an almost three-fold risk of being subjected to any of the factors indicating increased risk during labour and delivery [aOR 2.99(1.87-4.79)]. More than 2.5 hours from admission to the following CTG, more than 2.5 hours between recordings, and CTG non-interpretable due to poor quality were all factors that became significantly more common in the later period (p-value <0.01).

Performance of fetal blood sample when indicated was still not performed in 45.9% of the cases, compared to 39.3% (ns), and uterine tachysystole occurred in 37.4% of the CTG-recordings vs. 34.1% in the previous review.

The risk of a complex instrumental delivery did, however, decrease significantly from 14.1% of the cases to 5.1% (p-value <0.01).

## 8 DISCUSSION

### 8.1 THE IMPLEMENTATION OF THE NATIONAL PERINATAL PATIENT SAFETY PROGRAMME

To properly understand the PPS programme and assess its potential outcomes there is a need to examine not only the different features of the programme but also the implementation process of this large scale change intervention. Theoretical models from the field of dissemination and implementation science can help us understand how and why various strategies to change the behaviour of health professionals succeed or fail, and with this knowledge we can design appropriate strategies to improve care and reduce errors and harm (86, 87). One of the renowned researchers within the field expresses it like this:

“Implementation principles are like gravity – they are always present and working whether they are used intentionally or not. Unfortunately, attempts to use evidence-based interventions or other innovations are made without a lot of attention to implementation principles. Consequently, 5-15% success rates are typical for interventions that rely on people interacting with other people (changing behaviour of practitioners who interact with intended recipients). This is very different from administering chemicals where the intervention (the composition of the pill or serum) stays the same no matter who administers it.” **Dean Fixsen, PhD, Senior Scientist, Co-Director, National Implementation Research Network University of North Carolina at Chapel Hill, US (personal communication).**

In other words, implementation is the combination of processes with the intention to get an evidence-based intervention into practical use within a setting and hence reduce inappropriate care. Implementation science seeks to study these processes.

*Innovation* in organisations such as healthcare, is defined by Greenhalgh et al., p. 582 (88) as “a novel set of behaviours, routines and ways of working that are directed at improving health outcomes, administrative efficiency, cost effectiveness, or users’ experience, and that are implemented by planned and co-ordinated actions”. The term is sometimes used interchangeably with the term evidence-based intervention, which in implementation research is a wide concept of treatments proven effective, as well as denoting the nature of programmes, processes, policies and guidelines. The spread of such innovations is a complex, social and adaptive process in a continuum from “let it happen” via “help it happen” to “make it happen”. Special attention needs to be paid to the *setting*, the environment where the implementation occurs, and the *context*, the combination of unique factors or circumstances that actively interact with the implementation.

It is also important to distinguish between *diffusion* - the passive spread of new evidence-based practice, *dissemination* - active approaches to spread, and *implementation* - active

approaches to integrate an innovation within an organisation (88). The passive diffusion of innovations relies largely on how peers themselves work out how to use new knowledge. In an active dissemination approach, evidence-based interventions are spread to a specified audience of clinical practitioners through a predefined pathway using a planned approach (89). This can also be described as a 'push-and-pull-process' in which the recipients must want to accept the innovation (pull) and in a systematic way help to implement innovations (push). Implementation approaches rely on teams or individuals with management functions assuring that an innovation eventually becomes routine use in an organisation. During the last decades a growing acceptance of the need for research synthesis through meta-analyses has evolved, whereas the study of how to implement this evidence-based knowledge and practice is far less explored.

In this thesis the perceptions and effects of the three initial core interventions of the programme were studied in terms of the local teams' development of TMM and readiness for change (Paper I). The effects of the web-based CTG education programme were addressed (Paper II), and the hypothesis of pair-wise assessment of CTG compared to single assessment being more adequate was tested (Paper III). In Paper IV the local practices' improvement measures were described and outcomes in terms of the incidence of low Apgar score before and after the programme, and the changes in settled claims at LÖF over time were addressed. Finally, the risk of being subjected to situations indicating higher risk for low Apgar during labour and delivery, before and after the PPS programme was explored (Paper V). Thus, over time much information has been gathered that may shed light on the complexity of launching and evaluating large interventions such as the national PPS programme.

### **8.1.1 The PPS programme in the light of implementation science**

With the ambition to understand and guide implementation efforts, a number of theories and frameworks have been developed. One of these frameworks will be used to guide the discussion of different facets of the PPS and CTG programmes, combined with out-put and outcomes investigated.

#### CFIR, Consolidated Framework for Advancing Implementations Science

Based on an extensive review, the Consolidated Framework for Advancing Implementations Science (CFIR) was developed. CFIR is a synthesis of models, theories and frameworks used to "promote theory development and verification about what works where and why across multiple contexts" (90). Five major domains interact to influence the effectiveness of implementation: (I) the intervention characteristics; (II) inner setting; (III) outer setting; (IV) individuals involved; and (V) the implementation process.

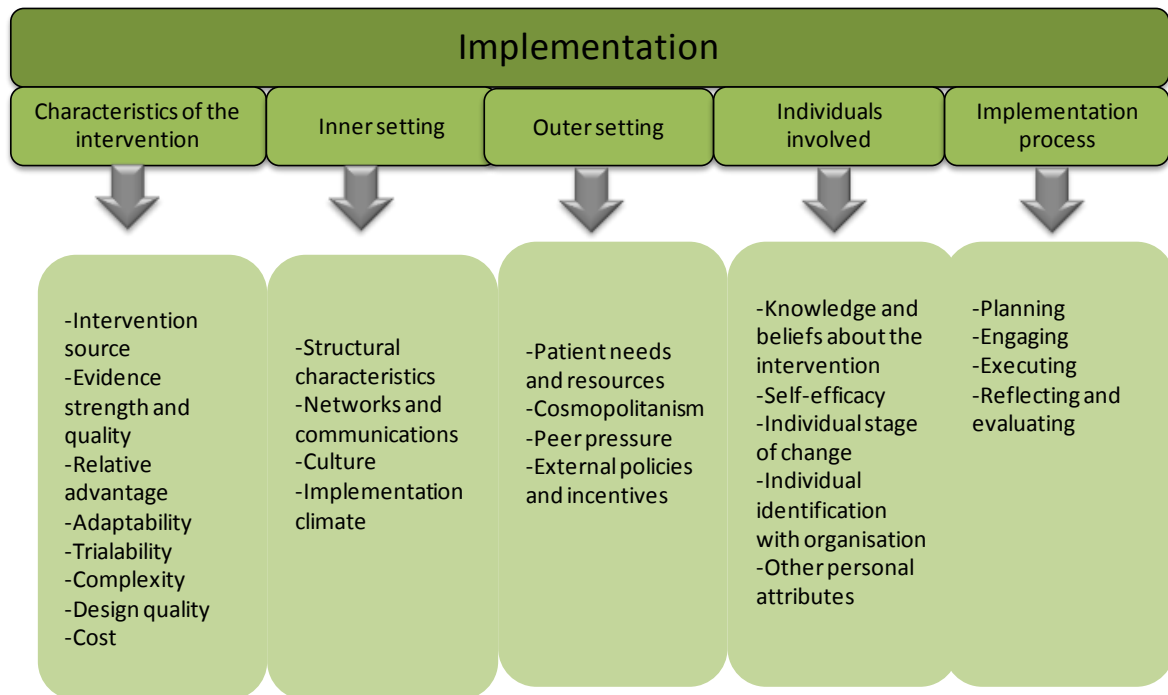


Figure 14. Implementation domains according to CFIR.

In this thesis, the CFIR domains considered relevant are used to address and discuss the results in Papers I-IV by focusing on elements that are important for a successful implementation of interventions such as the PPS and CTG programmes.

#### 8.1.1.1 Characteristics of the PPS and CTG interventions (I)

##### 8.1.1.1.1 Intervention source

An intervention will gain different legitimacy among the adopters depending on whether it starts off among frontline staff trying to solve a problem, or if it is developed by an external research group. The PPS programme was developed externally; however, it is likely that the positive response from all involved obstetric units (hereafter referred to as the units) was related to the general consensus among obstetricians and midwives about the goal, i.e. to avoid suboptimal care and improve outcomes for the newborn infant. The web-based CTG programme was a special feature within the PPS programme, a tailored education that was offered to all obstetric units free of charge. It was developed within the PPS programme and the two papers addressing this intervention also provided input to the CTG programme that was further developed and improved.

##### 8.1.1.1.2 Evidence strength and quality behind the interventions

Many implementation projects aim to apply evidence-based interventions. Perceived quality and validity of the evidence behind the intervention will have an impact, as well as

how these aspects are presented to health professionals. In the PPS programme, the evidence used to motivate change were gaps between evidence-based knowledge and current practice that were the results of the studies of patient injury claims and substandard care in Swedish settings (23, 37, 81). The main deficiencies were found in fetal surveillance and use of oxytocin and were associated with substandard care and preventable asphyxia in newborns. These findings were the benchmark for the professional organisations when designing the quality improvement project in co-operation with LÖF. The aim of the programme being an incentive to improve care for the newborns was, in that sense, deeply rooted within the group of clinically active members. The identification of risk areas forming the base for the questions in the self-assessment were identified in a process based on empirical clinical experience (Figure 5). During the evaluation process, it has become evident that, although these risk areas were in line with findings from other researchers, the subdivision into measurable entities was lacking and had to be constructed subsequently.

On a national level the programme contained four core elements offered by the steering committee to all participating units, i.e. the self-assessment, the peer review process, the mutual agreement for change, and the final report. A joint assembly for sharing experiences and lessons learned was offered to all participants after every sequence. The stakeholders (the management at the units) were presented with background data of gaps between knowledge and practice, benchmarking the initiative, the planned activities and facilitating functions at the initiating meeting at the unit.

The choice of self-assessment as a method for improvement work has strong support in the literature (91, 92). It was considered a useful tool to elucidate weaknesses and strengths in the processes of delivery care at the units. This helped the units to focus on areas in which improvements were needed. By involving multi-professional teams when discussing the questions, several perspectives on safety for the newborn were illuminated. During this cooperative process, new team mental models emerged. The self-assessment also served as a preparation for the following exchange of knowledge and experiences with the peer reviewers (Paper I).

The evidence of peer review as an intervention during implementation or organisational development is less studied and details about the supporting function in a change process are basically unknown (93). The peer review process was appreciated and perceived as an opportunity to share knowledge and confirm the ideas of the local teams. The discussions with the peer reviewers encouraged the teams to clarify their perceptions of the improvement efforts. Apparently, though, it had only limited impact on team mental models and did not add new ideas on how improvements should be achieved.

The mutual agreement for change is similarly unexplored by researchers but seems to have added some positive pressure on the units to perform explicit goals in the improvement work. Thus, the role of peer reviewing and an agreement for change to support an improvement process needs further research in order to be effectively used.

Previous standardised education programmes have shown positive results on staff's ability to classify CTG (94, 95). Mandatory training in fetal monitoring has resulted in decreasing incidence of low Apgar in clinical micro-systems (29, 96). When launching the programme there was reason to believe that an easily accessible, web-based education in fetal assessment would be considered a worthwhile initiative to improve knowledge among obstetric staff.

#### 8.1.1.1.3 Adaptability of the PPS and CTG programmes

There is a strain between the elements in an intervention that are considered non-exchangeable (core) and those factors that are adaptable for the local setting. The adaptable factors must be tailored to fit a certain setting if the involved health professionals are going to accept it. The structure of PPS programme with four core interventions, subsequently connected in a fixed time frame were, in that sense, not flexible, but the work processes of the teams and interpretation of the content was adaptable to local needs and pre-requisites. This highly organic and adaptive process in which the organisation adapts to the innovation and vice versa is a process that is not easily controlled by external change agencies.

The area of organisational risk was intrinsic in all features of the self-assessment tool and therefore a persisting component in almost all measures of improvement addressed by the units (Paper IV). Organisational risks are defined differently by researchers but can be viewed as the interacting junctions between administrators, policy makers and technology suppliers, and the practitioners meeting patients. These risks might be best managed by a combination of "putting in extra barriers, by questioning the notion of normal operations, by reducing complexity or by enhancing the control system or leadership commitments around it" (p.100) (97).

The intention of the first question in the self-assessment was to help the units to describe their organisational context, often referred to as the structure measures, described in 4.2.5. In the final reports there are many examples of improvement efforts directed at these structural measures e.g. up-dating existing or constructing new guidelines for known risk areas such as fetal assessment and neonatal resuscitation. We know less about their implementation and use in clinical practice and even less about their sustainability (Paper IV). It is likely that the question in the self-assessment, explicitly asking for guidelines and routines within the different risk areas, has directed the focus to this measure.

The second question in the self-assessment -“How do you make sure that these routines are followed?” is potentially helpful when addressing process measures and creating systems for assuring reliability of care processes (54, 98). The exploring of this question by the local teams, and later together with the peer reviewers was supposed to bring about insights about, for example, the need for systems to support regular follow-up of adherence to guidelines, or how to continuously secure high competence in fetal surveillance. In the final report, about half of the units described activities indicating that new process measures were adopted (Paper IV). Although this is yet to be formally studied, it is obvious when reading the self-assessments that a deeper knowledge of the implications of these measures was lacking. For example, to the question “How do you ensure that guidelines for the management of oxytocin are followed?” the reply from the local units showed a large variation. The answers ranged from “We aim at introducing regular random samples of medical records and follow the adherence”, indicating a plan for continuous follow-ups of the reliability of this specific care process. At the other range of the spectrum were answers like: “It is noted in the medical record”. These answers indicate that the question would have needed further understanding of the concept of measures to be adequately managed.

The peer review process was also a core intervention with a fixed work process. The peer reviewers followed a formalised protocol when scrutinising the self-assessment, planned explorative questions to examine possible improvements more thoroughly, and later compiled a feedback report. Adaptability was accomplished by the co-operation between the peer review group and the local teams in which the peer reviewers helped the managerial staff to prioritise the most important improvement measures and tailor those to fit the local needs, making it flexible. The units were encouraged to focus on improvements that they found most motivating at the time. However, this necessary adaptability also contributed to an under-specification, leading to a loss of ability to follow the process of implementation and explain and understand findings and outcomes (61).

When planning the intervention, different approaches were discussed. The reason for not choosing a standardised protocol-based process was mainly that neither the professional organisations nor LÖF have the mandate to execute a controlling function. The peer reviewers were encouraged to support the management in its effort to improve patient safety but had no authority to demand adherence to a pre-determined design. Knowledge about implementation theory and evaluation methods were also lacking at the time.

The CTG programme intervention was adaptable in the sense that all units were free to use it as they found most suitable for their local needs. Only 37% of the units reported that they had introduced it, and a mandatory certification was only planned in 17% (Paper IV). Another 17% expressed that they “urged employees to undertake the education” and one



of the 46 units reported that “the programme was launched”. Three obstetric units in Paper V reported that they planned to implement mandatory education with the CTG-programme and three units had no intention to do so.

This ambiguous approach to introducing mandatory training might be seen from a variety of angles. Interestingly, among safety-critical fields, healthcare stands out in its views on competence, i.e. once a clinician is authorised there are no or few demands for recurrent assessment of technical and non-technical skills (4). Neither are there systemised controls of the competence of clinicians in handling new tasks, new technology or ability to collaborate in a team. In other industries preoccupied with safety, these issues are managed with competence checking, regular re-certification, standard communication and set phraseology for difficult tasks (4). The reason for this might be found in a difference in basic assumptions. Physicians still struggle with the historical idea that they are unique craftspeople whose individual mastery and virtues hold the complex patchwork of healthcare together (99). At the same time, it is the individual that is in focus in cases of adverse events, ending up in a simultaneous belief in human strength and brittleness different from other high reliability organisations.

Perhaps getting buy-in from healthcare professionals in adopting a new education that will take time and effort and at the same time put them at risk of appearing unqualified is difficult. Interestingly, a mandatory web-based education in fetal surveillance, including skills checking and an examination test was adopted by the Danish obstetric professionals when launching a quality improvement programme for perinatal safety (<http://www.regioner.dk/sundhed/kvalitet/patientsikkerhed>). Health professionals failing the examination for the third time are not allowed to handle the classification of CTG. Obviously, other factors are involved in forming attitudes towards skills verification. CFIR points out that perception of the evidence supporting the use of a specified intervention by the targeted individuals or local teams strongly affects the implementation effectiveness. Evidence considered valid by an outside expert group may well be perceived entirely differently by the individuals in the organisations set to use it, a conflict illuminated by the great variation in the adopting frequency of the CTG programme.

#### 8.1.1.1.4 Trialability of the interventions

A pilot intervention of the PPS programme was carried out (see 4.3.8). It is important to evaluate usability of a planned intervention as this increases the success rate of its adaptation (100). Improvements of the self-assessment were made and the manual for the peer review protocol used to compile the feedback report was also clarified. Further improvement of the instructions to the peer reviewers were made following Sequence I, as some suggestions on improvements were perceived as impractical by the receiving teams at the units (Paper I).

#### 8.1.1.1.5 Complexity of the intervention process and the challenge of evaluation

The complexity of an intervention increases with the number of sub-processes included. From a national point of view, the PPS programme, including four sequenced, well-defined steps leading up to the final report, seems comprehensible. When considering the large number of targeted organisations and the diversity of adopters, from management to frontline staff, from small units with 200 deliveries annually, to large city hospitals with more than 7,000 deliveries a year, this complexity increases. External influencing factors such as available resources, size and contextual variation are confounding factors, restricting the understanding of the variation of measures and their application. It is also important to mention that improvements were reported by unit representatives, which limit the accuracy of the data.

There was great variation between top -and lower performing units in terms of organisational maturity for systematic improvement work, and structures for quality improvement tools. This is mirrored by the diversity of activities in the final reports. Some units put most of their efforts into establishing and updating guidelines whereas others with well-functioning guideline systems could focus on implementing new process measures. In ultra-safe organisations like aviation, it's been pointed out that the safety measures needed in a specific setting are related to the level of maturity that has already been achieved (101). Reliable structures and processes are the foundation of safety, and trying to implement features of high reliability too prematurely in a safety evolution process might even decrease safety (chapter 14) (6). This raises questions on whether it would be possible to assess the base line maturity of a unit when initiating an improvement programme and then to measure progress. It is possible that specific questions in the self-assessment could be used to evaluate this maturity at baseline. An assessment of changes in patient safety culture might also have aided later reflection and learning. Knowledge about methods to do so was not available when initiating the project but has developed since then and is probably useful when planning further improvement (61).

Yet another example of this complexity can be seen in the results of the CTG-education programme (Papers II and III). The baseline level of correctly classified CTG tracings before the web-based programme was launched in 2009 was 64%. In 2012 the corresponding figure was 75%. In one obstetric unit the baseline was even higher, 93%. It is possible that the evaluation in Paper II was made too early, as changes in modes of operations take time. It is likely that the programme had been used in 2009-2012 but details on how it was managed at the local units, whether it was mandatory or optional, has not been studied. However, these figures indicate that high levels of correctly classified CTGs are possible to achieve and that there are lessons to be learned from the units with high scores.

#### 8.1.1.1.6 Cost of the PPS CTG programmes

Avoidable injuries due to severe delivery related asphyxia accounts for more than 20% of the total indemnification cost for LÖF. The aim with the monetary incentive was to stimulate a process of organisational awareness of the local patient safety structures and continuous and sustainable improvement efforts. It was distributed twice during the programme. The first payment of approximately EUR 5,000 on entering the project and EUR 15,000 on the completion of the final report was the same for all units. The units were granted economic compensation for five days per peer reviewer that participated in the project. Together with arrangements and travelling logistics around the introduction meeting, start-up seminars and visits to the units, the costs for LÖF amounted to around EUR 30,000 per unit.

Costs for the units, allowing local teams to engage in the work process, as well as costs for implementing improvements, have not been calculated.

The CTG programme was also funded by LÖF, explaining how it could be offered free of charge to the units. The construction and maintenance up until today has cost approximately EUR 100,000 and it is continuously maintained for about EUR 10,000 annually. As pointed out before, units managed the education very differently, some committing several hours with full salary for all front-line staff to complete the education, some not adopting the intervention at all, making it impossible to estimate the costs per unit.

#### 8.1.1.2 Inner and outer settings including when implementing PPS and CTG programmes

The inner setting is characterised by the social architecture in which people interact. Well-functioning relationships between people and a good team spirit can positively influence implementation efforts. Organisational culture and climate must be addressed from a multitude of perspectives. Within this domain the readiness for implementation is considered in which leadership engagement is a major dimension (102). Other researchers also point out that senior management and commitment is needed for successful implementation, as well as a readiness for change and engaging frontline staff in improvement processes (61, 103). The introduction meeting was used to direct the commitment of the unit managers to the aims of the PPS programme. Perhaps involving higher levels of management, such as hospital management or even the County Council Director would have aided the project, by adding important input on the design of the programme and by establishing an even deeper anchoring within the settings.

A 'clinical champion' (nurse/midwife or doctor) spending time to implement the intervention, performing regular reviews and re-planning in response to changing

situations, is thought to aid the process (60). The representatives for the steering committee most likely acted as clinical champions when launching the programmes to stakeholders and many units probably had local clinical champions during the process. However, the role of leadership within PPS programme is yet to be studied.

The outer setting includes the degree of networking with external organisations and the co-operation with peers. The basic conditions varied greatly between units, considering the variation in size, location and demography. For the small units, well-functioning networks with neighbouring units with more resources are crucial. In larger organisations perhaps internal communication between the obstetric and neonatal staff and the anesthesiologists is the greater challenge. Inter-professional collaborations within and between hospitals with the aim of exchanging experiences, solve mutual challenges such as agreeing on the appropriate level of care for sick newborn, and learn from adverse events were initiated in 30% of the units (Figure 13). Access of highly competent obstetricians and neonatologist at all hours was an issue in many of the reports and difficulties of ensuring sufficient resources were brought up more often in hospitals situated in less populated areas of Sweden.

#### **8.1.1.3 Individuals involved**

Individuals interact dynamically with the organisation. The units engaged representatives from all front-line staff involved in delivery care in the change process. This included midwives, assisting nurses, obstetricians, neonatologists and anaesthesiologists. In order to achieve higher levels of learning and sustainable changes in patient safety, the involvement of both staff and managers is crucial. Sustainable changes assume changes in mental models, norms and culture.

An example of how the adoption of the CTG programme was influenced by individual beliefs and attitudes is the intense debate that evolved during the study period in Paper II on whether or not it is possible to classify a CTG tracing without the knowledge about the clinical context. This highlights the question about differentiating classification and action, nomenclature and expected interventions. Discussions arouse among pairs on whether a pathological CTG really should be classified as pathological if the woman was actively pushing. The comprehension when designing Paper II and III was that all tracings can be classified according to the guidelines by FIGO, actions on the other hand depend on clinical details. Ambiguities about the obstetric interventions based on specified CTG patterns and timeframes lead to disrupted communication and unclear expectations on actions for fetal resuscitation. Common nomenclature, illuminated as crucial for coherent understanding of CTG is what eventually can create consistent expectations on actions to optimise fetal outcome (25, 104).

#### 8.1.1.4 The implementation process of PPS and CTG programmes

The implementation process is commonly described around four major activities, i.e. planning, engaging, executing, and evaluating.

##### 8.1.1.4.1 Planning the national intervention

The planning of the PPS programme on a macro-level was made by the steering committee together with the programme administrator at LÖF. They set the time-plan for engaging the units in the pilot sequence and subsequently the rest of the units in three sequences. The timescale for all core activities was defined in order to facilitate completion. Less attention was paid to more structured evaluation of local effects.

##### 8.1.1.4.2 Engaging health professionals in the intervention

Engaging individuals with similar education and professional backgrounds is important to gain acceptance and authority for further implementation efforts. The professional organisations engaged experienced colleagues to build multi-professional peer review teams. It is reasonable to assume that the perception of the peer review process as very much appreciated was related to the fact that the peers were colleagues sharing similar realities and challenges in improving care in their own units. It is also possible that the exchange of experiences and knowledge that became obvious in Paper I was related to the fact that these colleagues functioned as peer opinion leaders with high credibility and representativeness. Members of the steering committee - being well-known within the networks of obstetricians, neonatologist and midwives - probably contributed to the willingness to participate by acting as expert opinion leaders.

##### 8.1.1.4.3 Executing the intervention process

The programme administrator also had the function of an external change agent, facilitating the process by arranging all start-up seminars for peer reviewers, introduction meetings with the local teams, consecutive meetings with the steering committee, keeping track of timeliness of task completion, both nationally and in contact with the units. This function is important as it helps execute the process.

##### 8.1.1.4.4 Reflecting on lessons learnt

Assemblies were held after each sequence aiming for reflection and learning through-out the process. In the final assembly in 2011, to which all participants over the three sequences were invited, 213 peer reviewers, local managers and local teams participated in a one-day seminar, sharing experiences and good examples. Expert groups were then engaged in specific safety issues that had evolved during the PPS programme. The web-based education in neonatal resuscitation ([www.neohlrutbildning.se](http://www.neohlrutbildning.se)) was the result of one of these groups that, together with the CTG programme is continuously updated and

maintained. A review of the extensive material from the 46 units, synthesising good examples together with current knowledge in selected risk areas is the foundation of advice on best practice and the basis for clinical guidelines, published on LÖF's website (<http://lof.se/patientsakerhet/vara-projekt/rekommendationer-och-rad/>). A final report on the experiences and learning points has also been published ([http://lof.se/wp-content/uploads/2015/05/slutrapport\\_saeker\\_foerlossning.pdf](http://lof.se/wp-content/uploads/2015/05/slutrapport_saeker_foerlossning.pdf)).

#### 8.1.1.4.5 Evaluating the effects of the programme

Evaluation of the programme was not planned in advance but developed over time as the programme was executed. It has been widely recognised that quality improvement collaborations are difficult to assess as they interact in complex, socio-technical environments that are constantly changing (61). Many levels of the organisations are involved, including executive management to front-line staff. Vincent points out that, for complex interventions that evolve over a longer time period and involve many hospitals, it is impossible to fully understand the impact of a large scale intervention (p. 375) (6).

A clinical microsystem is the staff, technology and care processes involved supporting a specific population of patients, for example an obstetric unit. It has been pointed out that the clinical microsystem might be the most meaningful level to approach with system changing interventions (105, 106). A recent example is seen in one of the Swedish, mid-size units, engaged in improvement work with the aim to decrease the rate of cesarean section among term nulliparas with a spontaneously starting delivery. A continuous focus on the chosen outcome measures (cesareans section in nulliparous women, umbilical cord pH <7 and Apgar score <4 at five minutes) in combination with organisational changes resulted in decreasing c-section rates in the targeted group. It is concluded that the combination of activities were effective in this particular setting but that it is unknown what part of the intervention could be considered core (107). When involving multiple microsystems, there is variation in implementation as well as in the cumulative effects of the intervention, and high-performing microsystem may be masked by the less successful performance of the majority (61). This is illustrated in Paper IV in which the units with the highest rate of Apgar score <7 at five minutes showed a significant decrease in Apgar score 4-6 at five minutes after the PPS programme whereas the opposite trend was seen among the units with the lowest rate of Apgar score <7. In addition, the heterogeneity in the design of improvement programmes, and the research methods used to assess them limits the evidence base (108-110). In fact, there are still no established advantages with any of the previously reviewed approaches to large-scale quality and safety programmes (111), and although the need for evaluative methods is well recognized there is no consensus on what methods to use (6).

#### 8.1.1.5 Conclusions of the implementation process

The scientific evidence, based on data from Swedish settings pointing out the need for improvements in obstetric care, assured that the aim of the project was relevant for clinically active members within the professional organisations. This deep anchoring was probably the reason for the positive response and participation rate of 100 % of the Swedish units. The financial and administrative support offered by LÖF and the programme administrator was a prerequisite for the realisation of the interventions.

The risk areas were identified based on clinical experience. The self-assessment questions were then drafted around these risk areas and appear useful in illuminating strengths and areas in need of improvement. The first question used was efficient in exposing structural measures in need of improvement. The second question had the potential to induce the development of process measures; however the units managed the questions of adherence to structural measures very differently, indicating a need for clarification of the concepts of measures.

External influencing factors such as available resources, size and contextual variation are confounding factors, restricting the understanding outcomes. When merging the result from many microsystems, high performers may be hidden by a larger number of less successful organisations. Moreover, the evidence supporting a particular design of improvement programmes or evaluating methods is lacking (108-110).

In retrospect, launching the CTG programme, relying completely on passive diffusion may not have been optimal as this resulted in a wide range of adopting patterns across the units. Common nomenclature, illuminated as crucial for coherent understanding of CTG is what eventually can create consistent expectations on actions to optimise fetal outcome (25, 104). However, a training programme cannot stand for itself. Univocal constructs in fetal assessment must be used in everyday work and thereby contribute to a cultural change in which assessment of CTG with coherent nomenclature becomes a part of work-as-done. It will then be possible to frame common expectations and specify processes according to CTG pattern and stage of labour.

There are lessons to be learnt from the implementation processes of the studied programmes. The core interventions were non-exchangeable, assuring that all units followed the same work process. Adaptability allowed the local teams to choose the safety issues that they found most relevant at the time. On the other hand, this adaptability also contributed to a weakness, illustrated by the loss of data in the evaluation. The under-specification of well-defined improvement measures makes assessment difficult, and local processes that were used to generate change, are unknown. Without the mandate to follow up the process of implementation by a standardised protocol-based process it is

difficult to evaluate results. This insight might be useful when planning similar initiatives in the future.

Involving senior leaders on all levels at an early stage might enhance the priority given to the project. Securing a variety of selected evidence-based process and outcome measures with high potential of benefits and low barriers improves the chances of success.

Measuring baseline performance and following performance data during programme implementation aids the interpretation of data. Engaging multi-professional teams in improvement cycles like PDSA to enhance change and learning and to identify barriers in the local setting are other examples of ways to improve the impact of programmes such as the PPS programme (112). Finally, committing time and resources, as well as involving expertise, in planning and executing continuous evaluation is crucial. Then we can learn even more about what works where and why.



## **8.2 CLINICAL GUIDELINES (TO STANDARDISE OR NOT TO STANDARDISE – THAT IS THE QUESTION)**

Improving guidelines is the measure that gained most attention of all, engaging 83% of the units and measuring the adherence to these guidelines where planned in almost half of these.

Standardized procedures are indeed a corner-stone of safe practice. Frontline staff contributes to safety by following the rules. However, it is also true that safety is created by intelligent and flexible individuals who know when to adapt regulations to the changing and unforeseeable reality.

The question “Why don’t people just follow the guidelines?” deserves some attention. As discussed in 4.1.2, the most intuitive explanation is the view on certain individuals as more accident prone. It might be true that some people tend to take bigger risks than others but this can’t explain systemised violation to safety standards. Another perspective is the deviance to normalisation, characterised by frequent, seemingly invisible violations to standard procedure that occurs so often and pass without anyone raising objections or even noticing that they are eventually built in to the system (113). Both midwives and physicians gradually adapt to this migration of the collective idea of safe adherence to guidelines. This is likely to be at least part of the explanation for the results in Paper V, indicating that uterine tachysystole is accepted for more than 20 minutes in more than one-third of the cases and oxytocin was increased in 40% of cases with pathological CTG, despite clear evidence for increased risks (Paper V, table 3). It might also be an explanation for the acceptance of the deviance from the standardised nomenclature when assessing CTG, in many units “not so good” or “the pattern is a bit so-so” is widely accepted.

In a review Cabana et al (114) describe one way of understanding barriers to guideline adherence by organising them into three major categories according to how they affect health practitioners’ knowledge, attitudes and behaviour, see figure 15. “Lack of awareness” in which the health practitioner does not know that the guideline exists, and “lack of familiarity”, when guideline content is insufficiently assimilated both affect the individuals’ knowledge. “Lack of agreement”, for example, due to disagreement on how evidence should be interpreted or that the guideline is oversimplified, also affects attitudes. Knowledge and attitudes together influences the behaviour that in turn depends on external factors like lack of time or resource.

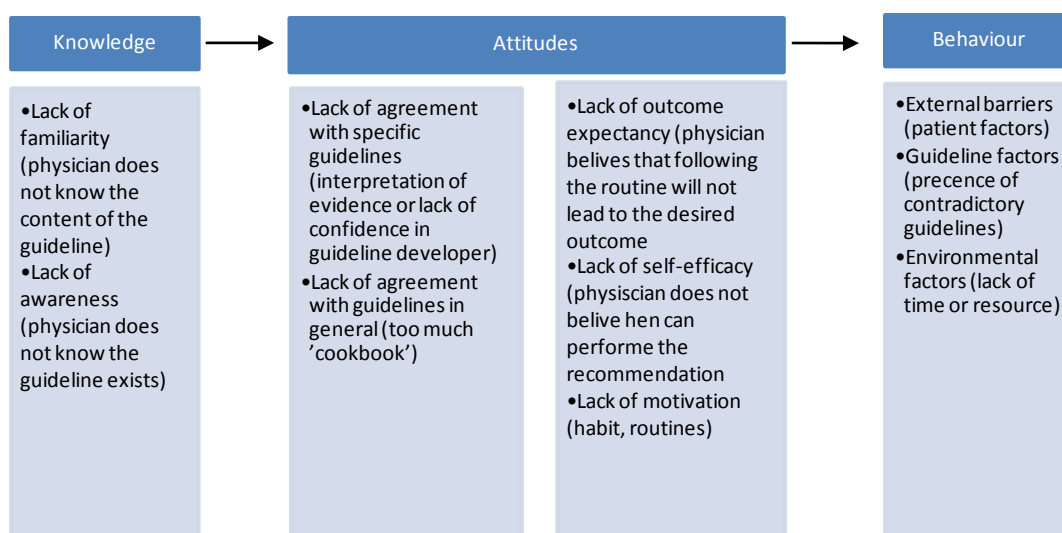


Figure 15. A framework for understanding barriers to adherence to guidelines. Adapted from Gurses et al, JAMA, October 20, 1999-Vol 282, No 15.

Gurses et al (115) highlight ambiguity as a major reason for healthcare professionals not to follow guidelines. Sometimes this ambiguity is due to lack of evidence, and yet the clinician is forced to intervene on the basis of the same unclear evidence.

Some variables in the definition of suboptimal care in Paper V are regulated in evidence-based guidelines such as the use of oxytocin, whereas others are more clinical based, such as the use of fetal blood sample. It is possible that different improvement approaches apply to different problems in the care processes. For oxytocin management the evidence base underpinning the guidelines might make it supportable with strict adherence checks continuously used as process measures. Fetal blood sample, on the other hand, is a tool considered helpful but the knowledge base that would support a more stringent use – and thereby monitoring against an expected level of reliability - is lacking.

Guidelines are often too extensively written and at the same time they often fail to provide guidance on how to prioritise what to do, when, how and by whom (115). Having a clear team mental model of practice guidelines improves patient outcomes (48, 115).

The number of policies recommending interventions at a specified condition can be extensive, making it time-consuming and difficult to find the most accurate advice. Often more than one version exists and sometimes with headings that are not logical. For example, is oxytocin guidelines best sorted under 'Augmented labour' or 'Oxytocin management' or even 'Dystocia treatment'? The length of a guideline document can also

impair the navigation to the most relevant information. Thus, there is a need for updated and improved guidelines as well as the systems used to organise them, not least in obstetrics. It has been suggested to use an application of knowledge from the area of human factors engineering when elaborating on guidelines, in which healthcare staff need to be involved to assure usability in clinical practice (116).

Finally, in order to understand how health professionals try to adapt to the contradicting goals of safety, productivity and constrained resources, it's helpful to look at Amalbertis' framework for violation and migration (fig 16) (117, 118). Although primarily developed for high reliability organisations like aviation and railway, the concepts are useful to understand how a system can move from safe operation to occasional disaster. The 'legal space' is the situation when a unit is working according to plan with not too many unexpected events to deal with and where procedures and processes are followed. Now add some pressure, a sudden increase in patient inflow, lack of staffing, and reduced resources. Staff adapts to get the job done and meet the needs of the patients and themselves by violating routines and stretching the boundaries for safe operations. As disasters are so very unusual, especially in obstetrics, and the advantages for both the health professional and the management are obvious, the system slowly accepts the migration from safe to the 'illegal normal' state. Once the system has accommodated this situation, any further stress to the system or individuals might easily result in an accident. When looking back to such an event, the violation of routines or guidelines by an individual health professional will be the most evident explanation. To detect migration of the whole system is very difficult for the untrained eye. Involvement of expert knowledge and science is essential to develop improved safety for patients.

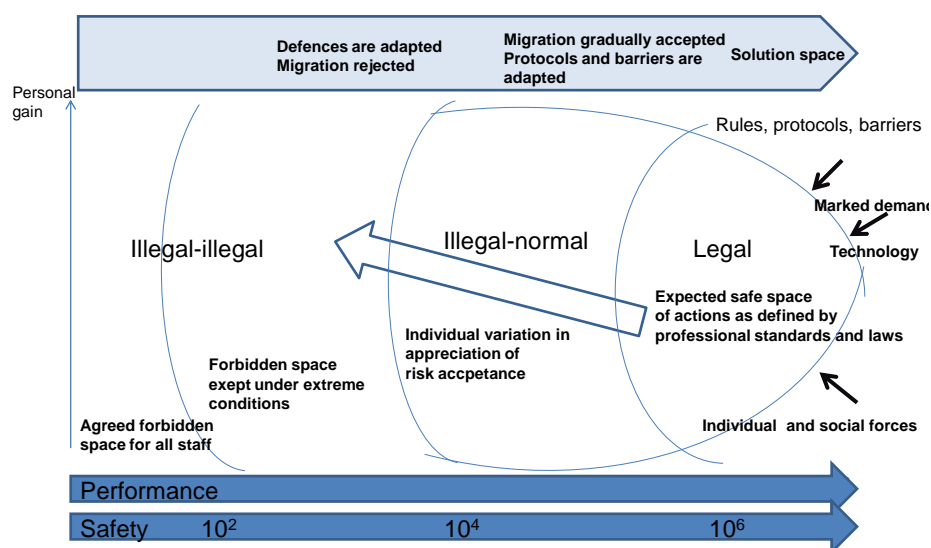


Figure 16. Adapted from Amalberti, R ; Vincent, C ; Auroy, Y ; de Saint Maurice, G, Violations and migrations in healthcare; a framework for understanding and management. 15 no.suppl\_1, (66-71) 2006

### 8.3 METHODOLOGICAL DISCUSSION AND CONSIDERATIONS

The major strength of this thesis is the comprehensive approach to explore effects and results of a collective effort to improve safety for the newborn. The use of mixed methods to illuminate different perspectives of measures helped us to learn and eventually understand more about the challenges of evaluating large scale initiatives. The major limitation of this evaluation is that data on the local processes of the PPS programme are unavailable. We have data about the input, we have evaluated the core interventions and we have explored the outputs and outcomes (final reports, Apgar, settled claims and suboptimal care) but the process in between is lost in the 'black box'.

#### 8.3.1 Qualitative data

##### 8.3.1.1 Most important strengths

Paper I was a process assessment of the first three core elements: self-assessment, peer review, and mutual agreement for change. The aim was to elucidate the output of the specified intervention initiating the PPS programme on the local level. The results can be seen as a measurement of the stakeholder's perception of advantages of these essential steps. Practically all eligible participants from the units were interviewed, reaching 99 % of the managers of the units as well as the head midwives and one senior consultant. The directed content analysis (a priori coding) can be helpful to focus the research questions

when there is usable knowledge or theory about a phenomenon such as TMM. The knowledge can then guide the discussion of the results and eventually, existing theories can be enhanced (84) (Paper I).

Final reports were collected from all units, in Sweden. This made it possible to compile how measures of improvement were managed locally and extract information on a national level. Qualitative data can provide important information on processes and reasons for certain outcomes (Paper IV).

#### **8.3.1.2 Most important limitations**

The interviews in sequence I of the PPS programme lasted between 15 and 30 minutes each, limiting the amount of data collected. It is possible that longer interviews at several occasions during the programme could have provided more details on how the TMM developed over time and how the initial views on the safety situation translated into actual actions and change. Interviews were also carried out, despite the fact that seven of the 14 units in sequence I had just recently, or had not at all, received the feedback report. This was due to prolonged peer review processes in some units and might have influenced the perception of this core intervention by the adopters. It's also important to mention that the instructions to the peer reviewers were slightly changed after sequence I, and peer reviewers were urged to take a humble approach when co-operating with local teams, remembering that not all units have the same prerequisites.

The directed content analysis has some limitations. Researchers already familiar with a theory about expected outcomes may be biased and more likely to find evidence that confirms the theory. When conducting the interviews, the researcher might ask questions directing answers in a particular way (84). To prevent important information being lost, an extra category for important information was added (Paper I).

The open-ended questions in the self-assessment resulted in a large variety of the structure, process and outcome measures. Knowledge about a structured framework for these measures (structure, process and outcome) was not available when creating the questions. Taking structure measures as an example, it is possible that the units chose to report what they had, to a greater extent than what they did not have in place. For process measures, the questions were often misunderstood, as general knowledge about this construct is low. Using predefined measures would have aided the evaluation and increased understanding of the improvement efforts. Improvement measures taken were reported by the units themselves which limit the accuracy of the result. Conducting observations of clinical practice and documenting actual adherence and adaptation of the interventions provides more reliable data but requires both time and resources (Paper IV).

### **8.3.2 Quantitative data**

#### **8.3.2.1 Most important strengths**

The randomisation procedure from the CTG pool was made to make sure that the observed proportion of correctly classified CTGs' is the same as would be observed in large samples if every participant had interpreted all 40 traces (Papers II and III). To quantify the statistical uncertainty, the risk that the two proportions (single and pairwise) differed due to pure chance, a confidence interval was calculated. The 95% CI for the proportion of correctly individually classified CTGs was 0.72-0.79. The lower limit is well above what is reported in former studies supporting the conclusion that there was a high standard of knowledge in these study populations (Paper III).

The quarterly cumulative incidence of settled claims for reported avoidable birth injuries saw a slightly decreased trend (0.5%;  $p=0.279$ ) from 2000 to 2012. This measure decreased further in 2012–2014 (7.5%;  $p=0.049$ ) (Paper IV). Patient injury claims rates are a source of patient safety data less subjected to bias than other patient generated data (119). In Sweden, data are aggregated on a national level, making it possible to identify unusual events. Moreover, claims are scrutinised by specialists, using well-defined criteria for the assessment of avoidability. Although underreporting can be assumed, there is also a bias towards more serious events, indicating that severe asphyxia afflicting newborns would have a high reporting frequency (Paper IV)

#### **8.3.2.2 Most important limitations**

The 'reference standard' CTG -pool used in Papers II and III was created from the 40 out of 55 intrapartum recordings that were classified with 100% inter-individual agreement by four experts, also involved in the creation of the web-based CTG education programme. The proportion of agreement was 73%. It is possible that other experts would have classified these recordings differently, resulting in a different 'standard reference'. Interestingly, in Paper II the proportion of agreement between the individuals in one of the maternity unit and the 'standard reference' pool was as high as 93%, indicating that high proportions of agreement are indeed possible to achieve (Papers II and III).

The power analysis in Paper II was based on previous studies reporting proportion of agreement on intermediate and abnormal tracings of 49-69 % (33, 34, 120) among experienced clinical staff, and 55% in an unselected group. Thus, the baseline level of knowledge at 64% correctly classified CTGs before the education was higher than expected. In Paper III the baseline level was even higher (75%). Perhaps a design involving more participants would have increased the chance of detecting an improvement after the education (Paper II).

#### 8.3.2.2.1 Apgar score and pH-data as outcome measures in large scale interventions

No difference in the incidence of low Apgar score was seen after the PPS programme. It is likely that a significant change of a single outcome like Apgar score was very difficult to influence, especially as the baseline level of performance was already so high (Figure 17). A further limitation is the retrospective, before- and after design that doesn't allow us to draw confident conclusions on causal correlations between the interventions and the results.

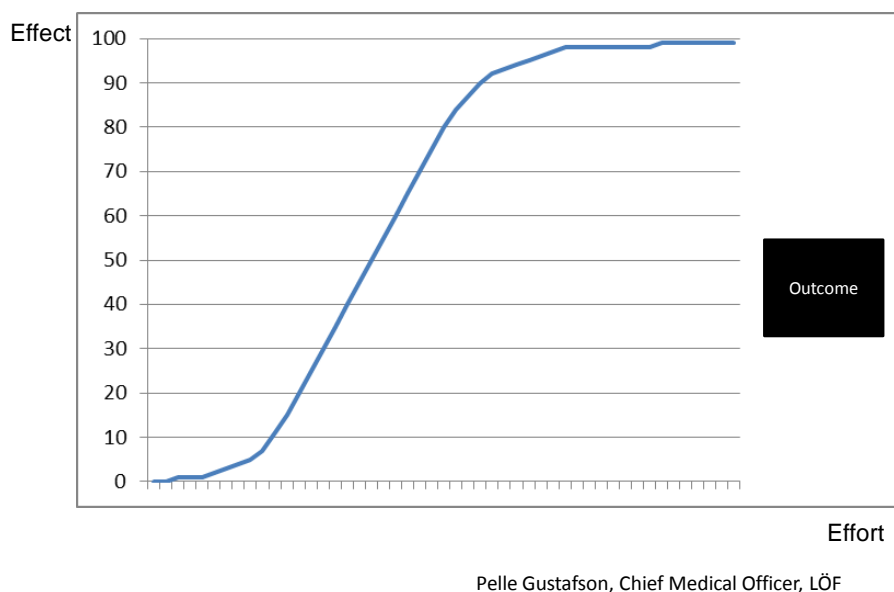


Figure 17. With high levels of performance, great efforts are needed to accomplish small changes and those small changes are difficult to detect (personal communication, Pelle Gustafson, Chief Medical Officer, LÖF)

The primary endpoint in Paper IV was Apgar <4 at five minutes, Apgar <7 at five minutes and/or low pH defined as <7.10 or 7.05. For the diagnosis of acute intrapartum asphyxia, a combination of Apgar score and umbilical cord pH is needed. While Apgar is performed in almost 100% of all deliveries, pH is still unreliably reported, restricting the possibility to analyse data that, if added, might have provided a more comprehensive picture. In settings with a low frequency of umbilical blood gas samples it is sometimes argued that tests are taken when the newborn shows signs of asphyxia. However, background data in Paper V showed that in the settings with the lowest frequency of pH-testing, umbilical blood gas samples were taken in only half of the cases with low Apgar score (data not shown). It is likely that routine testing of all umbilical cords at birth would provide the highest frequency of samples in stressed situations with an asphyxiated newborn. Until there are better alternative outcome data, more complete reports of pH-data in combination with Apgar score form the basis for local settings to continuously measure and monitor these

outcomes. It would also make it possible to aggregate national outcome data and discuss alternative methods, such as indexing as in AOI. Finally, the possibility for parents to discover the cause of injury is important. A single outcome measure reflecting quality of care during delivery is probably a utopia. Quality of care and improved safety systems might best be assessed with a combination outcome measures.

There are examples of successful interventions in perinatal settings. One example is the state-wide collaboration for perinatal patient safety in Michigan, a multi-centre intervention involving 15 hospitals during a period of 11 months (121). In this case, a significant improvement in patient safety culture and care processes was reported. Clark et al focused on the decreasing numbers of claims of litigation after redesigning safety processes in a large healthcare system of 120 facilities, where a focus on uniform guidelines and processes and an effective peer review process were reported as important components of success (27). There are also a number of single site interventions. Pettker et al introduced a multi-step intervention at a tertiary-level hospital in Yale, New Haven, including: an outside expert review; the establishment of a patient safety committee and a patient safety function (nurse); protocol standardisation; training in team skills and fetal heart monitoring interpretation (28). The intervention significantly reduced the AOI and improved safety climate (66). Wagner et al implemented a similar multi-component safety initiative at a tertiary-level hospital with 5,300 deliveries a year, resulting in a significant reduction in modified AOI. It is interesting to note that the individual safety marker Apgar <7 at five minutes did not change in any of these interventions (26).

#### 8.3.2.2.2 Settled claims

The decreased number of severely asphyxiated infants granted economic compensation by LÖF has to be interpreted with some caution. The main reason is the reporting system; only patients and relatives can file a claim even if the non-punitive system might lower the barriers for healthcare professionals to help patients with this process. Thus, the decrease in settled claims seen after the PPS programme can be an effect simply of a decreasing willingness to file a claim, or a tendency to file a claim later, i.e. after 2012 rather than before, factors that cannot be controlled for. This is unlikely, as it would mean that parents of an injured child would constitute an isolated group with a decreasing willingness to file a claim or to file a claim later; this in a situation where the number of filed claims to LÖF increase generally and tend to be made earlier. The lower number of settled claims could also be an effect of something other than the PPS programme, e.g. improved maternity care or better neonatal care. The increasing use of cooling treatment implemented from 2007 is likely to have affected the panorama of injuries to asphyxiated newborns and detailed knowledge about this is still lacking.



#### 8.3.2.2.3 Substandard care

The definition of substandard care in obstetrics differs depending on source. Most researchers use a definition including violation to guidelines and/or generally accepted management according to evidence or expert opinion (37, 122). The synonymous concept suboptimal care is also used in perinatal care (123).

The study reported in Paper V is a criterion-based case record review that attempts to explore errors as well as harm in obstetric care. The specific criterion were chosen as they were considered objective measures reflecting consensus based quality of care regulated in national guidelines. Explicit review has been shown to have low inter-rater variability (124).

Missing data about the increase of oxytocin when CTG was pathological, and uninterpretable CTGs' makes reliable conclusions difficult. In the present study the case control study from 2004 to 2006 was used as a comparison and we chose to compare cases with cases and controls with controls assuming that controls were representative for the population. The fact that only 0.3% of controls after the PPS programme were delivered instrumentally compared to 14.1% before may question that assumption. However, the overall rate of instrumental deliveries is between 7-9% in Stockholm and many of these infants could be expected to have Apgar <10 at five minutes.

It has been shown that knowledge of adverse neonatal outcome affects the overall classification of CTG when interpreting tracings retrospectively, especially affecting the judgment of decelerations and variability (125). We tried to overcome this potential hindsight bias by following the strict criterion-based protocol when collecting data and by performing an inter-rater variability test in CTG interpretation. The outcome bias is well known to influence the judgment of others in case reviews and root cause analyses. Interestingly, not only is the judgment of others more harsh but the willingness to make a judgment also increases after severe outcomes (126). The hindsight bias, on the other hand, is related to the reviewers' estimation of the predictability of an outcome when an event is scrutinised retrospectively ("I would have known that this was going to happen if I had been in charge") (127). Is it really possible to judge whether a team around a labouring woman did right or wrong by looking backwards in time? Isn't it impossible to collect all the crucial information that was actually available to the team at the time? These are the key questions of outcome and hindsight bias, and more research is needed to understand the mechanisms of these biases. It is also important to evaluate possible de-biasing strategies if clinical review and root cause/event analysis is going to be useful as a tool for learning and improving.

Substandard care is a term appearing to assume that there is a fixed construct that defines 'standard care'. Obviously there is not. Nevertheless, the term is used by many attempting

to measure safety in perinatal care. In my opinion, the concept of substandard care is unfortunate as it implies that caregivers have been careless. It is likely that it provokes associations to recklessness, incompetence and negligence among staff at the sharp end. It is all too well known how naming, shaming and blaming has been part of culture in healthcare for a long time and that ending up with the fool's cap is a risk we take as front-line staff. This is probably not the best seedbed for a desire to learn about and improve care that we all struggle to deliver in the very best and safest way, every day and every night.

In perinatal care, adverse events are prevented by skilled, flexible and fore-seeing front-line staff on a daily basis. To learn from our every-day work is one of our major tasks. The challenge is to integrate and improve the well-established but limiting Safety I perspective- of which root-cause and event analysis are examples- with the Safety II mindset in which safety is not only measured by the rare adverse events but also by the many situations that go right. Efforts must be made to learn from the resilience that we already possess and then develop these capacities.

## 9 CONCLUSIONS AND CLINICAL IMPLICATIONS

Based on the results on the five papers some conclusions can be drawn:

- Self-assessment can be a useful tool when exploring strengths and areas of possible improvement in a healthcare unit. New team mental models about patient safety improvement emerged during the work process. Although there were indications of the peer review process being confirmative and the mutual agreement adding positive pressure for change, further studies are needed to fully understand how peer reviewing can be used to enhance the improvement efforts (Paper I).
- There were no measurable improvements in the ability to classify CTG patterns in accordance with a 'reference standard' after education and examination in the web-based CTG programme. The baseline knowledge in the population was higher than expected (Paper II).
- No significant difference was seen between single interpreters and couples working together. However, the baseline of consensus with the 'reference standard' was even higher in this study (75% among individuals and 80% among pairs) and one unit had a consensus level of 93% with the 'reference standard' (Paper III)
- Numerous improvement efforts were made in all Swedish units, mainly focusing on structural measures. Half of the units introduced new process and/or outcome measures with the intention to continuously measure adherence to clinical guidelines and outcomes. The incidence of low Apgar was unchanged during the study period but the number of settled claims due to severe delivery-related asphyxia showed a decreasing trend (Paper IV).
- The overall risks for suboptimal care increased twice among controls and three times among cases in 2009-2012 compared to 2004-2006. Improving guidelines is important but not enough to change practice among front-line staff. Reliable methodology to evaluate and assess quality of care is in need of further exploration (Paper V).

## 9.1 PLANNING FOR FUTURE IMPROVEMENT INTERVENTIONS

During the research process, new insights concerning methodological issues involved in planning, executing and evaluating large-scale improvement programmes and patient safety has evolved, raising new questions.

Studies are needed to understand which processes in perinatal care that are suitable for standardisation and reliability measures and which processes that should be approached from a Safety II direction, acknowledging the complexity and unpredictability that is the reality of obstetric care.

Agreement on process measures with high relevance for obstetric care needs to be developed. With basic standards, procedures and structures for continuous maintenance, the next step towards more reliable care processes can be taken. A possible approach would be to evaluate a standardised management of key clinical protocols such as oxytocin and fetal surveillance. The aim would be to identify barriers and facilitators to guideline adherence in perinatal settings.

We need to explore which outcome measures that can reflect quality of care in obstetrics. Outcome measures should be unambiguous with low estimated numbers of unknown cases in order to be reliable. Delphi processes have successfully been used in other obstetric settings and could be used to confirm the concepts of outcomes and induce a discussion about alternative constructs, such as indexing as in AOI. We also need to evaluate if lactate is a useful outcome measure.

Studies are needed to explore if it is possible to define different levels of organisational maturity concerning patient safety within healthcare. Further, as structure measures were successfully identified by the self-assessment, it could be evaluated if self-assessments could be used to elucidate the specific needs of an organisation so that further improvement efforts can be tailored to fit the evolutionary process of the individual unit.

Exploration of the features of resilience in perinatal settings is needed. We need to draft studies that can help us understand what clinicians do to adjust to changing conditions. What are the tools of an obstetrician and midwife, subsequently evaluating an escalating situation during labour? What are the obstacles to resolve an unexpected situation? Which conditions support flexibility for frontline staff?

Studies are needed to explore if the register 'Graviditetsregistret' ('Pregnancy register') will be a useful and reliable source for outcomes, as it collects nationwide medical data about pregnancies and deliveries. ([www.medscinet.com/gr/mal.aspx](http://www.medscinet.com/gr/mal.aspx)).

Hindsight bias is a major barrier in the current methods used to learn from adverse events, such as root-cause-and event-analysis. Research is needed to understand the impact and increase knowledge about de-biasing strategies.

Finally, future research is suggested to develop the understanding of migration of an organisational system and to find ways to control for this risk.

## 10 ACKNOWLEDGEMENTS

First of all I would like to express my gratitude to my supervisors for everlasting patience, encouragement and guidance!

Charlotta Grunewald, my main supervisor. Thank you for inviting me into the world of obstetrics and gynecology in December 1997. It was my passion and you made it possible for me to realise that dream. From that day on, you have been my role model and my coach. You have taught me accuracy and high standards, with high expectation on me, and I think, even more on yourself. I have learned so much from you, such as “everything can always become better”. Whenever I have needed your help, you have been there for me. We have had so much fun! But most of all, you have become my dear friend.

Monica Nyström, my co-supervisor. Thank you for opening the door to a parallel reality, that of implementation science. Despite a constantly busy calendar, you have always made me feel prioritized. Your vibrant and fascinating explanations on input, output and black boxes, painting rectangles and arrows, trying to make the rest of us understand what you see so clearly has been truly inspiring, educating and enjoyable.

Ulf Högberg, my co-supervisor. Thank you for saying just the right things when I needed them most. Your support and encouragement has been crucial to me. Thank you for enriching this work and our studies with your humble and warm attitude, wisdom and deep knowledge about science.

I would also like to thank:

Karin Pukk Härenstam, for being a true mentor and with your brilliance, knowledge and experience, helping me to get a grip of this process.

Hans Pettersson, statistician at KI SÖS, for endless patience through hours and hours of statistical coaching and teaching.

Sissel Saltvedt, former head of the obstetric section at Södersjukhuset, for backing me up through my doctoral studies and for always finding a way to say something up-lifting, spreading good spirit and delight around you.

Margareta Hammarström. You told me to think twice before entering this project. You said that it would be difficult, complex and very challenging. I wish I could have shared with you that you were right. And thank you for giving me full support when I decided to go for it anyway.

Sophie Berglund, for sharing your passion and knowledge about patient safety and research.

Torbjörn Schultz, Jon Ahlberg and Pelle Gustafsson, all at the Swedish National Patient Insurance Company LÖF, for stimulating discussions about patient safety improvement and for academic contribution to my research as well as instant and qualified help with the administrative details on the project. Your ambition to constantly improve safety for the patients is admirable. Jonas Ranstam, thank you for statistical contribution and insights.

Ingegerd Lantz, head of the Department of Obstetrics and Gynecology and Eva Östlund head of the obstetric section for making it possible for me to proceed with my doctoral studies and for always being encouraging and supportive.

My wonderful colleges and midwives at the Women's Department at Södersjukhuset, thank you for being so smart, creative, and energising. You are all the reason why I love to go to work.

Stina Wirsén, illustrator, for your generous contribution with the cover picture, I adore your drawings!

Marion Lindh and Magna Andreen Sachs, for inviting me into the world of health improvement and patient safety.

Anders Dahlström, colleague, role model and friend. Your passion for CEPS, education, and communication inspires me every day. You always remind what the most important issue is- our patients! You fight for improved safety with an including, encouraging attitude that invites the rest of us to grow. You are a true clinical champion.

CEPS-Obstetrics colleagues; David, Emma, Jenny, Anna, Magdalena, Cathrine, Anna-Karin, Maie and Gerty. With you the ambition to improve care, communication and teamwork becomes real. We have had times of doubt and times of success. But most of all, we have had times of great fun. Let's never stop!

Patrik Lundberg, for putting me in contact with the people that became the most influential to my personal journey into patient safety, Johan Bergström and Sidney Dekker. Karin Sartorius and Patrik together, for being wonderful friends throughout all phases of life.

Johan Bergström och Sidney Dekker, for opening my eyes to a new world, a new language, a new mind-set. The course in Lund was the turning point for me, pointing out a completely new direction. I must admit I was shaken but I loved every second.

Lena Molander, for helping me with the data collection, Magdalena Cruz for support of all kinds and, Crippa Marsnäs, Ebba Insulander, Laura Araya Oyarzo and Monica Berggren for always helping me getting in touch with cyber space, quickly, capably and kindly.

Jeanette Öhrman and Anne Edgren, for helping me through the administrative jungle of the doctoral studies.

David Gustafsson; for putting yourself at my disposal, guiding me through the trickiest journey of them all.

My lovely friends, I'm so lucky to have all of you around me, your phone calls and texts, your encouragement and all the times we spend together. Thank you all for reminding me of what really matters- relations, love and sharing each others' life.

My family, Inger, Maria, Ann-Charlotte, Rasmus and Fabian for love and care in the life we share. My mother Kina, for your endless love and for always believing that I could do whatever I wanted to. I wish my father Reje could have been here with us, he shed so much love and light upon us. My dear brother Fredrik, my sister-in-law Anette, Nellie, Emil and Julia, my supporting and joyful companions in life.

Joachim, my love. you always make me laugh, no matter what and you put things in the right perspective. You are my solid ground and my challenge. Thank you for your care, support, patience and love.

Alice, and Felix, my children, with you everything is obvious. You are the meaning of life. You fill me with energy and deep gratefulness. I love you so much.

The studies in this thesis were supported by grants from Swedish National Patient Insurance Company LÖF, SFP-vård, Karolinska Institutet and Södersjukhuset.



## 11 SAMMANFATTNING PÅ SVENSKA

Förlossningsvårdens utmaning är att försöka förebygga skador och agera i tid vid tecken till avvikelse men samtidigt undvika interventioner i det normala förloppet. Att barn skadas så allvarligt i samband med födelsen att det leder till livslångt funktionshinder eller för tidig död är ovanligt. I Sverige beräknas c:a 35 barn per år födas med svåra syrebristskador som uppstått under förlossningen. Det exakta antalet barn som får syrebristskador till följd av felbehandling i anslutning till förlossning är okänt, men årligen får 20-50 barn med svår förlossningsrelaterad skada ersättning från Patientförsäkringen LÖF (LÖF) då skadan bedömts som undvikbar.

Patientsäkerhet är ett viktigt fundament i en hälso- och sjukvård av god kvalitet. De bakomliggande orsakerna när säkerheten brister rör ofta organisationen och dess rutiner. En process som bygger på förändring av hela organisationen snarare än på bestraffning av enskilda individer ökar möjligheten att åstadkomma varaktiga förbättringar i patientsäkerheten (128).

Vetenskaplig kunskapsutveckling är central för att förbättra vårdens insatser. Mätning av resultat och effekter förbises ofta i skadepreventiva åtgärdsprogram. Vetenskaplig evidens saknas för värdet av en nationell, övergripande intervention för ökad patientsäkerhet i förlossningsvården. Det vetenskapliga underlaget för detta avhandlingsarbete är studier baserade på Patientförsäkringen LÖF:s journalarkiv över förlossningsrelaterad syrebrist eller tidig död. De vanligaste orsakerna till felbehandling var bristande fosterövervakning, försummelse av tecken på syrebrist och felbehandling i anslutning till förlossningen (23). Två tredjedelar av barn födda inom Stockholms Läns Landsting med tecken på syrebrist under åren 2004-2006 och en tredjedel av friska kontroller under samma tidsperiod utsattes för suboptimal vård (37). Det konstaterades att majoriteten av de faktorer som orsakar suboptimal vård borde vara möjliga att förebygga genom utbildning och ökad medvetenhet om risker. Mot bakgrund av dessa resultat pågick under åren 2008-2011 en nationell intervention, Projekt Säker Förlossningsvård (PSF), med det övergripande målet att förbättra patientsäkerheten inom förlossningsvården för att minska antalet barn med undvikbara förlossningsrelaterade skador. PSF byggde på tvärprofessionell samverkan och omfattade Sveriges alla 46 förlossningskliniker. Projektet bestod av en process med följande innehåll: lokal självvärdering, kollegial granskning utförd av ett multiprofessionellt revisorsteam, överenskommelse om förbättringsåtgärder och en skriftlig slutrapport om de åtgärder som genomförts på de lokala förlossningsklinikerna. För att möta behovet av ökad kunskap om fosterövervakning utvecklades inom ramen för projektet ett webbaserat utbildningsprogram i fosterövervakning och CTG-tolkning som tillhandahölls gratis till alla förlossningskliniker i Sverige. Detta möjliggjordes genom LÖF som också bidrog med ekonomiskt och administrativt stöd under hela projektet.

I Studie I genomfördes 80 telefonintervjuer med en klinikchef, en chefsbarnmorska och en förlossningsöverläkare på 27 förlossningskliniker i Sverige. Syfte var att belysa hur interventionen, utförd i en given ordning, påverkade teamen och deras "mentala modeller" av patientsäkerhetsarbete under interventionsprocessen. Resultatet visade att självvärderingen ansågs vara värdefull och fungerade som ett användbart instrument för att belysa styrkor och identifiera förbättringsområden på respektive förlossningsklinik. Den kollegiala granskningen var uppskattad trots att den inte ansågs ha så stor påverkan på teamets arbetsprocess. Den skriftliga feedbacken från revisorsteamet och den gemensamma överenskommelsen om åtgärder ansågs fungera som stimulerande påtryckning för förändringar i verksamheten.

I studie II undersöktes om förmågan att tolka CTG (hjärtljudsregistrering) förbättrades efter genomgång och examination i ett datoriserat CTG-utbildningsprogram. En pool av referens-CTG-kurvor skapades av de 40/55 CTG kurvor som tolkades med 100 % samstämmighet av en grupp expertbedömare. Dessa 40 CTG-kurvor användes sedan som "facit" i studie II och III. Målgruppens förmåga att klassificera CTG jämfördes för varje individ, före och efter genomgången utbildning. Den undersökta populationen bestod av 179 barnmorskor och läkare på en av landets största förlossningskliniker. I den vetenskapliga utvärderingen sågs ingen signifikant förbättring i förmågan att klassificera en CTG-kurva i enlighet med "facit" efter genomgången utbildning. Kunskapsnivån före utbildningen låg högre än förväntat. Vid subgruppsanalys sågs en förbättring i förmågan att klassificera normala CTG-kurvor samt att andelen rätt klassificerade kurvor ökade när bedömningen gjordes parvis.

I studie III undersöktes enskild jämfört med parvis tolkning av CTG. 536 barnmorskor och läkare tolkade först en slumpvis utvald CTG-kurva individuellt. Därefter tolkades en ny CTG-kurva i par med en kollega (268 par). Ingen signifikant skillnad i förmågan att tolka CTG i enlighet med "facit" kunde ses mellan individuell och parvis tolkning. Även här låg den basala kunskapsnivån högre än väntat. Intressant nog fanns det relativt stora skillnader mellan den basala kunskapsnivån på de olika klinikerna. Personalen på av de studerade klinikerna tolkade CTG i enlighet med "facit" i 93 % av fallen jämfört med 68 % på en annan klinik.

I studie IV användes tre olika metoder för att belysa nationella effekter av PSF. Dels analyserades texten i de skriftliga slutrapporterna och de förbättringsåtgärder som genomförts på klinikerna sammanställdes. Dels jämfördes andelen barn i Sverige som föddes med tecken till syrebrist vid förlossningen, före, under och efter PSF. Dessutom analyserades antalet barn med svår förlossningsrelaterad syrebrist som sökt och fått ersättning från LÖF. Resultatet visar att 83 % av förlossningsklinikerna rapporterade att man uppdaterat och förbättrat riktlinjer och rutiner och att c:a hälften av klinikerna infört

nya mätmetoden för att följa upp följsamhet till riktlinjer och utfallsmått, så som tecken på syrebrist. Andelen barn i landet som föddes med tecken till syrebrist förändrades inte över tid (2000-2012). Däremot sågs en vikande trend bland det antal barn som fick ekonomisk ersättning från LÖF.

I studie V studerades den faktiska handläggningen under förlossning och jämfördes med i förväg uppställda kriterier. Dessa kriterier beskrev detaljer för vad som anses vara evidensbaserad övervakning av fostret under förlossning, hantering av värkstimulerande läkemedel och handläggning vid komplicerad instrumentell förlossning som t.ex. sugklocka. De fall som studerades var barn som föddes med tecken till syrebrist under åren 2009-2012 och för varje fall studerades en kontroll (ett barn som föddes utan tecken till syrebrist). Journaldata från mödravård och förlossning, inklusive CTG och neonatalvård analyserades och överfördes till ett protokoll som använts i en tidigare studie med identiskt design. Som jämförelse användes data från en fall-kontroll-studie av barn födda i Stockholm 2004-2006 (37). Resultatet visar att övervakning av barnet under förlossning snarast försämrats under den senare perioden, för både fall och kontroller. Bland fallen ökade risken för att utsättas för oförsiktig användning av värkstimulerande medel men risken för traumatisk instrumentell förlossning minskade över tid.

## **11.1 SLUTSATSER OCH ÖVERVÄGANDEN**

Sammanfattningsvis har ett omfattande förbättringsarbete ägt rum på landets alla förlossningskliniker. Våra resultat visar dock att det fortfarande finns områden inom förlossningsvården som behöver förbättras. Att utvärdera resultat av mångfacetterade och resurskrävande projekt är avgörande för att kunna mäta nyttan av insatserna och samtidigt förstå vilka faktorer som underlättar och vilka som försvårar förbättringsarbetet. En utmaning är att stora projekt innefattar många skilda aktörer som alla är verksamma i en ständigt föränderlig miljö. Genom att redan vid inledningen av ett projekt definiera ett urval av evidensbaserade process- och utfallsmått som lämpar sig för standardisering ökar chanserna att man ska lyckas följa upp resultaten av projektet. Först då blir det möjligt att kontinuerligt mäta följsamhet och tillförlitlighet i vårdprocesser. Att uppdatera riktlinjer och rutiner är viktigt men det räcker inte. Samtidigt behöver vi utveckla förståelse för den resiliens som redan finns i vårdorganisationen och bredda kunskapen om hur vi utvecklar dessa förmågor. Kunskap om fosterövervakning, standardiserad nomenklatur och åtgärder vid hotande syrebrist behöver utvecklas.

## 12 REFERENCES

1. Leape LL, Brennan TA, Laird N, Lawthers AG, Localio AR, Barnes BA, et al. The Nature of Adverse Events in Hospitalized-Patients - Results of the Harvard Medical-Practice Study-II. *New England Journal of Medicine*. 1991;324(6):377-84.
2. Donaldson L. An organisation with a memory. *Clinical Medicine*. 2002;2(5):452-7.
3. Reason JT. Individ-och systemmodeller för felhandlingar- att skapa rätt balans i hälso-och sjukvården. In: Ödegård S, editor. *Patientsäkerhet Teori och praktik*. Första upplagan. Stockholm: Liber; 2013.
4. Dekker S. *Patient safety: a human factors approach*. 1 ed. Boca Raton: CRC Press; 2011.
5. Cook R. Systemperspektiv på säkerhet. In: Ödegård S, editor. *Patientsäkerhet Teori och praktik*. Första upplagan. Stockholm: Liber; 2013.
6. Vincent C. *Patient Safety*. 2nd ed. UK: Wiley-Blackwell; 2006 2010.
7. Reason J. *Human Error*. 1 ed. UK: Cambridge University Press; 1990.
8. Reason JT. *Managing the risks of organizational accidents*. 1 ed. Aldershot: Ashgate; 1997.
9. Hollnagel E, Woods, D.D. & Leveson, N. *Resilience engineering Concepts and Precepts*. 1 ed. Aldershot: Ashgate; 2006.
10. Woods DD, Dekker, S, Cook, R, Joannessen, L, Sarter, N. *Behind Human Error*. Second edition ed. Burlington: Ashgate; 2010.
11. Cook R, Woods, D, Miller, C. *A Tale of Two Stories; Contrasting Views of Patient Safety*. PDF. US National Patient Safety Foundation, 1998.
12. Reason JT. Human error: models and management. *BMJ (Clinical research ed)*. 2000;320(7237):768-70.
13. Rutberg H, Grunewald, Charlotta, Spencer, Anna. Patientsäkerhetsarbetet i Sverige har gett resultat. *Lakartidningen*. 2015;112:DEXI.
14. Act TSPS. [http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Patientsakerhetslag-2010659\\_sfs-2010-659/?bet=2010:659](http://www.riksdagen.se/sv/Dokument-Lagar/Lagar/Svenskforfattningssamling/Patientsakerhetslag-2010659_sfs-2010-659/?bet=2010:659). 2016.
15. Rasmussen J. Risk management in a dynamic society: a modelling problem. *Safety Science*. 1997;27(2-3):183-213.
16. Weick KE, Sutcliffe, K.M. *Managing the Unexpected: Resilient Performance in an Age of Uncertainty* 2ed: Blackwell Publishing Inc; 2009.
17. Reason JT. Understanding adverse events: the human factor. In: Vincent C, editor. *Clinical Risk Management: Enhancing Patient Safety*. 2nd ed. London: BMJ Books; 2001.
18. Vincent C. *Understanding how things go wrong. Patient Safety*. UK: BMJ Books; 2010. p. 141-67.

19. Wrigstad J, Bergström J, Gustafson P. Mind the gap between recommendation and implementation—principles and lessons in the aftermath of incident investigations: a semi-quantitative and qualitative study of factors leading to the successful implementation of recommendations. *BMJ Open*. 2014;4(5).
20. Lundberg J, Rollenhagen C, Hollnagel E. What-You-Look-For-Is-What-You-Find – The consequences of underlying accident models in eight accident investigation manuals. *Safety Science*. 2009;47(10):1297-311.
21. Lundberg J, Rollenhagen C, Hollnagel E. What you find is not always what you fix—How other aspects than causes of accidents decide recommendations for remedial actions. *Accident Analysis & Prevention*. 2010;42(6):2132-9.
22. Hollnagel E, Braithwaite, J. *Resilient Health Care*. Farnham, GB: Ashgate; 2013.
23. Berglund S, Grunewald C, Pettersson H, Cnattingius S. Severe asphyxia due to delivery-related malpractice in Sweden 1990-2005. *BJOG : an international journal of obstetrics and gynaecology*. 2008;115(3):316-23.
24. Steer PJ. Surveillance during labour. *J Perinat Med*. 2009;37(5):451-6.
25. Knox GE, Simpson KR. Perinatal high reliability. *American journal of obstetrics and gynecology*. 2011;204(5):373-7.
26. Wagner B, Meirowitz N, Shah J, Nanda D, Reggio L, Cohen P, et al. Comprehensive perinatal safety initiative to reduce adverse obstetric events. *Journal for healthcare quality: official publication of the National Association for Healthcare Quality*. 2012;34(1):6-15.
27. Clark SL, Belfort MA, Byrum SL, Meyers JA, Perlin JB. Improved outcomes, fewer cesarean deliveries, and reduced litigation: results of a new paradigm in patient safety. *American journal of obstetrics and gynecology*. 2008;199(2):105.e1-7.
28. Pettker CM, Thung SF, Norwitz ER, Buhimschi CS, Raab CA, Copel JA, et al. Impact of a comprehensive patient safety strategy on obstetric adverse events. *American journal of obstetrics and gynecology*. 2009;200(5):492.e1-.e8.
29. Draycott T, Sibanda T, Owen L, Akande V, Winter C, Reading S, et al. Does training in obstetric emergencies improve neonatal outcome? *BJOG : an international journal of obstetrics and gynaecology*. 2006;113(2):177-82.
30. Thorngren-Jerneck K, Herbst A. Low 5-minute Apgar score: a population-based register study of 1 million term births. *Obstetrics and gynecology*. 2001;98(1):65-70.
31. Commission TJ. Sentinel event alert issue 30--July 21, 2004. Preventing infant death and injury during delivery. *Advances in neonatal care : official journal of the National Association of Neonatal Nurses*. 2004;4(4):180-1.
32. Santo S, Ayres-de-Campos D. Human factors affecting the interpretation of fetal heart rate tracings: an update. *Current opinion in obstetrics & gynecology*. 2012;24(2):84-8.
33. Westerhuis M, van Horen E, Kwee A, van der Tweel I, Visser GHA, Moons KGM. Inter- and intra-observer agreement of intrapartum ST analysis of the fetal electrocardiogram in women monitored by STAN. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2009;116(4):545-51.
34. Ojala K, Makikallio K, Haapsamo M, Ijas H, Tekay A. Interobserver agreement in the assessment of intrapartum automated fetal electrocardiography in singleton pregnancies. *Acta obstetrica et gynecologica Scandinavica*. 2008;87(5):536-40.

35. Ayres-de-Campos D, Bernardes J, Costa-Pereira A, Pereira-Leite L. Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. *British journal of obstetrics and gynaecology*. 1999;106(12):1307-10.
36. Belfort MA, Saade GR, Thom E, Blackwell SC, Reddy UM, Thorp JMJ, et al. A Randomized Trial of Intrapartum Fetal ECG ST-Segment Analysis. *New England Journal of Medicine*. 2015;373(7):632-41.
37. Berglund S, Pettersson H, Cnattingius S, Grunewald C. How often is a low Apgar score the result of substandard care during labour? *BJOG: an international journal of obstetrics and gynaecology*. 2010;117(8):968-78.
38. Jonsson M, Norden-Lindeberg S, Ostlund I, Hanson U. Metabolic acidosis at birth and suboptimal care--illustration of the gap between knowledge and clinical practice. *BJOG : an international journal of obstetrics and gynaecology*. 2009;116(11):1453-60.
39. Pehrson C, Sorensen JL, Amer-Wahlin I. Evaluation and impact of cardiotocography training programmes: a systematic review. *BJOG: an international journal of obstetrics and gynaecology*. 2011;118(8):926-35.
40. Institute for Safe Medication Practices. High-alert medications. Huntington Valley, Philadelphia, 2007. Available from [www.ismp.org/survey/Survey200702W.asp](http://www.ismp.org/survey/Survey200702W.asp)
41. Jonsson M, Norden-Lindeberg S, Ostlund I, Hanson U. Acidemia at birth, related to obstetric characteristics and to oxytocin use, during the last two hours of labor. *Acta obstetricia et gynecologica Scandinavica*. 2008;87(7):745-50.
42. Scalise D. Clinical communication and patient safety. *Hospitals & health networks / AHA*. 2006;80(8):49-54, 2.
43. Lewis. Saving mothers' lives: reviewing maternal deaths to make motherhood safer – 2003–2005. The Confidential Enquiry into Maternal and Child Health (CEMACH). The Seventh Report of the Confidential Enquiries into Maternal Deaths in the United Kingdom. London: 2007 Contract No7.
44. Dekker SW. Accidents are normal and human error does not exist: a new look at the creation of occupational safety. *International journal of occupational safety and ergonomics: JOSE*. 2003;9(2):211-8.
45. Siassakos D, Crofts JF, Winter C, Weiner CP, Draycott TJ. The active components of effective training in obstetric emergencies. *BJOG: an international journal of obstetrics and gynaecology*. 2009;116(8):1028-32.
46. Siassakos D, Fox R, Bristowe K, Angouri J, Hambly H, Robson L, et al. What makes maternity teams effective and safe? Lessons from a series of research on teamwork, leadership and team training. *Acta obstetricia et gynecologica Scandinavica*. 2013;92(11):1239-43.
47. Jeffery AB, Maes, J.D, Bratton-Jeffery, M.F. Improving team decision-making performance with collaborative modeling. *Team Performance Management: An International Journal*. 2005;11:40-50.
48. Hysong SJ, Best RG, Pugh JA, Moore FI. Not of One Mind: Mental Models of Clinical Practice Guidelines in the Veterans Health Administration. *Health Services Research*. 2005;40(3):829-48.
49. Davison G, Blackman D. The role of mental models in innovative teams. *European Journal of Innovation Management*. 2005;8(4):409-23.

50. Forsner T, Wistedt AA, Brommels M, Janszky I, de Leon AP, Forsell Y. Supported local implementation of clinical guidelines in psychiatry: a two-year follow-up. *Implementation Science*. 2010;5.
51. Woods DD, Cook IR. Nine Steps to Move Forward from Error. *Cognition, Technology & Work*.4(2):137-44.
52. Donabedian A. Evaluating the quality of medical care. 1966. *Milbank Q*. 2005;83(4):691-729.
53. Pronovost PJ, Nolan T, Zeger S, Miller M, Rubin H. How can clinicians measure safety and quality in acute care? *Lancet*. 2004;363(9414):1061-7.
54. Simpson KR. Measuring perinatal patient safety: review of current methods. *Journal of obstetric, gynecologic, and neonatal nursing: JOGNN / NAACOG*. 2006;35(3):432-42.
55. Draycott T, Sibanda T, Laxton C, Winter C, Mahmood T, Fox R. Quality improvement demands quality measurement. *BJOG: an international journal of obstetrics and gynaecology*. 2010;117(13):1571-4.
56. Boulkedid R, Alberti C, Sibony O. Quality indicator development and implementation in maternity units. *Best practice & research Clinical obstetrics & gynaecology*. 2013;27(4):609-19.
57. Sibanda T, Fox R, Draycott TJ, Mahmood T, Richmond D, Simms RA. Intrapartum care quality indicators: a systematic approach for achieving consensus. *European journal of obstetrics, gynecology, and reproductive biology*. 2013;166(1):23-9.
58. Simms RA, Yelland A, Ping H, Beringer AJ, Draycott TJ, Fox R. Using data and quality monitoring to enhance maternity outcomes: a qualitative study of risk managers' perspectives. *BMJ Qual Saf* 2014;23:457-464
59. Mann S, Pratt S, Gluck P, Nielsen P, Risser D, Greenberg P, et al. Assessing quality obstetrical care: development of standardized measures. *Joint Commission journal on quality and patient safety / Joint Commission Resources*. 2006;32(9):497-505.
60. Ovretveit J, Klazinga N. Learning from large-scale quality improvement through comparisons. *International journal for quality in health care : journal of the International Society for Quality in Health Care / ISQua*. 2012;24(5):463-9.
61. Benn J, Burnett S, Parand A, Pinto A, Iskander S, Vincent C. Studying large-scale programmes to improve patient safety in whole care systems: challenges for research. *Social science & medicine (1982)*. 2009;69(12):1767-76.
62. Simpson KR. Quality improvement in perinatal care. *MCN The American journal of maternal child nursing*. 2011;36(5):277.
63. Bax M, Nelson KB. Birth Asphyxia: a Statement. *Developmental Medicine & Child Neurology*. 1993;35(11):1022-4.
64. Martin JA, Hamilton BE, Sutton PD, Ventura SJ, Menacker F, Kirmeyer S, et al. Births: final data for 2005. *National vital statistics reports : from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*. 2007;56(6):1-103.
65. Hogan L, Ingemarsson I, Thorngren-Jerneck K, Herbst A. How often is a low 5-min Apgar score in term newborns due to asphyxia? *European journal of obstetrics, gynecology, and reproductive biology*. 2007;130(2):169-75.

66. Pettker CM, Thung SF, Raab CA, Donohue KP, Copel JA, Lockwood CJ, et al. A comprehensive obstetrics patient safety program improves safety climate and culture. *American journal of obstetrics and gynecology*. 2011;204(3):216.e1-6.
67. Thorngren-Jerneck K, Herbst A. Perinatal factors associated with cerebral palsy in children born in Sweden. *Obstetrics and gynecology*. 2006;108(6):1499-505.
68. Moster D, Lie RT, Irgens LM, Bjerkedal T, Markestad T. The association of Apgar score with subsequent death and cerebral palsy: A population-based study in term infants. *The Journal of pediatrics*. 2001;138(6):798-803.
69. ACOG Committee Opinion. Number 333, May 2006 (replaces No. 174, July 1996): The Apgar score. *Obstetrics and gynecology*. 2006;107(5):1209-12.
70. Stuart A, Otterblad Olausson P, Kallen K. Apgar scores at 5 minutes after birth in relation to school performance at 16 years of age. *Obstetrics and gynecology*. 2011;118(2 Pt 1):201-8.
71. Burrell C. Cerebral palsy: intrapartum causal criteria used in courts. *BJOG: An International Journal of Obstetrics & Gynaecology*. 2014;121(13):1728-.
72. Yeh P, Emary K, Impey L. The relationship between umbilical cord arterial pH and serious adverse neonatal outcome: analysis of 51,519 consecutive validated samples. *BJOG : an international journal of obstetrics and gynaecology*. 2012;119(7):824-31.
73. Badawi N, Kurinczuk JJ, Keogh JM, Alessandri LM, O'Sullivan F, Burton PR, et al. Antepartum risk factors for newborn encephalopathy: the Western Australian case-control study. *BMJ (Clinical research ed)*. 1998;317(7172):1549-53.
74. Holzmann M, Nordstrom L. Follow-up national survey (Sweden) of routines for intrapartum fetal surveillance. *Acta obstetrica et gynecologica Scandinavica*. 2010;89(5):712-4.
75. Heinis AMF, Spaanderman ME, Gunnewiek JMTK, Lotgering FK. Scalp blood lactate for intra-partum assessment of fetal metabolic acidosis. *Acta obstetrica et gynecologica Scandinavica*. 2011;90(10):1107-14.
76. Sibanda T, Sibanda N, Siassakos D, Sivananthan S, Robinson Z, Winter C, et al. Prospective evaluation of a continuous monitoring and quality-improvement system for reducing adverse neonatal outcomes. *American journal of obstetrics and gynecology*. 2009;201(5):480.e1-6.
77. Pagani F. Peer review as a tool for cooperation and change. *African Security Review*. 2002;11(4):15-24.
78. Kristensen B. The Impact of Quality Monitoring on Institutions: a Danish experience at the Copenhagen Business School, Quality in Higher Education. 1997;3:1:87-94.
79. Berglund S. Severe Asphyxia due to Substandard Care During Labour 2010. [Thesis for doctoral degree]. Available from <https://openarchive.ki.se/xmlui/handle/10616/1>
80. Berglund S, Grunewald C, Pettersson H, Cnattingius S. Risk factors for asphyxia associated with substandard care during labor. *Acta obstetrica et gynecologica Scandinavica*. 2010;89(1):39-48.
81. Berglund S, Norman M, Grunewald C, Pettersson H, Cnattingius S. Neonatal resuscitation after severe asphyxia-a critical evaluation of 177 Swedish cases. *Acta paediatrica*. 2008;97(6):714-9.



82. Amer-Wohlin I, Herbst, A. Fosterövervakning under förlossningen. Statens Beredning för Medicinsk och Social Utvärdering. Sweden 2008.
83. Nordström L, Waldenström, U. State of the Art: Handläggning av normal förlossning (Handling of Normal Labour). Sweden: Socialstyrelsen, 2001, 010206.
84. Hsieh H-F, Shannon SE. Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*. 2005;15(9):1277-88.
85. Pastor R, Guallar E. Use of two-segmented logistic regression to estimate change-points in epidemiological studies. *American Journal of Epidemiology*. 1998;148(7):631.
86. Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N. Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *Journal of clinical epidemiology*. 2005;58:107-12.
87. Wallin L. Knowledge translation and implementation research in nursing. *International Journal of Nursing Studies*. 2009;46(4):576-87.
88. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q*. 2004;82(4):581-629.
89. Rabin BA, Brownson RC, Haire-Joshu D, Kreuter MW, Weaver NL. A glossary for dissemination and implementation research in health. *Journal of Public Health Management and Practice*. 2008;14(2):117-23.
90. Damschroder L, Aron D, Keith R, Kirsh S, Alexander J, Lowery J. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation science: IS*. 2009;4:50.
91. Samuelsson P, Nilsson LE. Self-assessment practices in large organisations: Experiences from using the EFQM excellence model. *International Journal of Quality & Reliability Management*. 2002;19(1):10-23.
92. Zink KJ, Schmidt A. Practice and implementation of self-assessment. *International Journal of Quality Science*. 1998;3(2):147-70.
93. Russell J, Armitage J. Peer review effectiveness: An analysis of potential loopholes within the USA Peer Review Program. *Managerial Auditing Journal*. 2006;21(1):46-62.
94. Blix E, Oian P. Interobserver agreements in assessing 549 labor admission tests after a standardized training program. *Acta Obstet Gynecol Scand*. 2005;84(11):1087-92.
95. Beckley S, Stenhouse E, Greene K. The development and evaluation of a computer-assisted teaching programme for intrapartum fetal monitoring. *BJOG*. 2000;107(9):1138-44.
96. Katsuragi S, Ikeda T, Noda S, Onishi J, Ikenoue T, Parer JT. Immediate newborn outcome and mode of delivery: use of standardized fetal heart rate pattern management. *The journal of maternal-fetal & neonatal medicine: the official journal of the European Association of Perinatal Medicine, the Federation of Asia and Oceania Perinatal Societies, the International Society of Perinatal Obstet*. 2013;26(1):71-4.
97. Dekker S. Patient Safety: A Human Factors Approach. 1 ed. New York: CRC Press, 2011.

98. Berenholtz SM, Pronovost PJ. Monitoring Patient Safety. *Critical Care Clinics*. 2007;23(3):659-73.
99. Amalberti R, Auroy Y, Berwick D, Barach P. Five system barriers to achieving ultrasafe health care. *Annals of internal medicine*. 2005;142(9):756-64.
100. Rycroft-Malone J, Kitson A, Harvey G, McCormack B, Seers K, Titchen A, et al. Ingredients for change: revisiting a conceptual framework. *Quality & safety in health care*. 2002;11(2):174-80.
101. Amalberti R. The paradoxes of almost totally safe transportation systems. *Safety Science*. 2001;37(2-3):109-26.
102. Weiner BJ, Amick H, Lee SYD. Conceptualization and measurement of organizational readiness for change - A review of the literature in health services research and other fields. *Med Care Res Rev*. 2008;65(4):379-436.
103. Vaughn T, Koepke M, Kroch E, Lehrman W, Sinha S, Levey S. Engagement of Leadership in Quality Improvement Initiatives: Executive Quality Improvement Survey Results. *Journal of Patient Safety*. 2006;2(1):2-9.
104. ACOG Practice Bulletin No. 106: Intrapartum fetal heart rate monitoring: nomenclature, interpretation, and general management principles. *Obstetrics and gynecology*. 2009;114(1):192-202.
105. Barach P, Johnson JK. Understanding the complexity of redesigning care around the clinical microsystem. *Quality & safety in health care*. 2006;15 Suppl 1:i10-6.
106. Nelson E, Batalden PB, Godfrey, MM. *Quality By Design: A Clinical Microsystems Approach*: San Fransisco, Jossey-Bass; 2007.
107. Blomberg M. Avoiding the first cesarean section—results of structured organizational and cultural changes. *Acta obstetricia et gynecologica Scandinavica*. 2016; Vol.95(5), pp.580-586
108. Schouten LM, Hulscher ME, van Everdingen JJ, Huijsman R, Grol RP. Evidence for the impact of quality improvement collaboratives: systematic review. *BMJ* 2008;336(7659):1491-4.
109. Ovretveit J, Gustafson D. Evaluation of quality improvement programmes. *Quality & safety in health care*. 2002;11(3):270-5.
110. Bate P, Mendel P, Robert G. *Organizing for Quality: The Improvement Journeys Of Leading Hospitals in Europe and United States*. United Kingdom: Radcliff Publishing; 2008.
111. Grol R, Baker R, Moss F. Quality improvement research: understanding the science of change in health care. *Quality & safety in health care*. 2002;11(2):110-1.
112. Langley GJ, Moen R, Nolan KM, Nolan TW, Norman CL, Provost LP. *The improvement guide: a practical approach to enhancing organizational performance*: John Wiley & Sons; 2009.
113. Vaughan D. Theorizing Disaster: Analogy, historical ethnography, and the Challenger accident. *Ethnography*. 2004;5(3):315-47.
114. Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PA, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. *JAMA : the journal of the American Medical Association*. 1999;282(15):1458-65.
115. Gurses AP, Seidl KL, Vaidya V, Bochicchio G, Harris AD, Hebden J, et al. Systems ambiguity and guideline compliance: a qualitative study of how intensive care

units follow evidence-based guidelines to reduce healthcare-associated infections. *Quality and Safety in Health Care*. 2008;17(5):351-9.

116. Carthey J, Walker S, Deelchand V, Vincent C, Griffiths WH. Breaking the rules: understanding non-compliance with policies and guidelines 2011-09-13 14:40:35.

117. Amalberti R, Vincent C, Auroy Y, de Saint Maurice G. Violations and migrations in health care: a framework for understanding and management. *Quality & safety in health care*. 2006;15:I66-I71.

118. Rasmussen J. Human Error and the Problem of Causality in Analysis of Accidents. *Philos T Roy Soc B*. 1990;327(1241):449-62.

119. Pukk Härenstam K. Learning from patient injury claims: an assessment of the potential of patient injury claims to a safety information system in healthcare [Thesis for Doctoral degree]. Stockholm: Karolinska institutet; 2007. Available from <https://openarchive.ki.se/xmlui/handle/10616/1>

120. Blix E, Oian P. Interobserver agreements in assessing 549 labor admission tests after a standardized training program. *Acta obstetrica et gynecologica Scandinavica*. 2005;84(11):1087-92.

121. Simpson KR, Knox GE, Martin M, George C, Watson SR. Michigan Health & Hospital Association Keystone Obstetrics: a statewide collaborative for perinatal patient safety in Michigan. *Joint Commission journal on quality and patient safety / Joint Commission Resources*. 2011;37(12):544-52.

122. Evers ACC, Brouwers HAA, Nikkels PGJ, Boon J, Van Egmond-Linden A, Groenendaal F, et al. Substandard care in delivery-related asphyxia among term infants: prospective cohort study. *Acta obstetrica et gynecologica Scandinavica*. 2013;92(1):85-93.

123. Jonsson M, Agren J, Norden-Lindeberg S, Ohlin A, Hanson U. Suboptimal care and metabolic acidemia is associated with neonatal encephalopathy but not with neonatal seizures alone: a population-based clinical audit. *Acta obstetrica et gynecologica Scandinavica*. 2014;93(5):477-82.

124. Graham WJ. Criterion-based clinical audit in obstetrics: bridging the quality gap? Best practice & research *Clinical obstetrics & gynaecology*. 2009;23(3):375-88.

125. Ayres-de-Campos D, Arteiro D, Costa-Santos C, Bernardes J. Knowledge of adverse neonatal outcome alters clinicians' interpretation of the intrapartum cardiotocograph. *BJOG : an international journal of obstetrics and gynaecology*. 2011;118(8):978-84.

126. Hugh TB, Dekker SW. Hindsight bias and outcome bias in the social construction of medical negligence: a review. *Journal of law and medicine*. 2009;16(5):846-57.

127. Fischhoff B. Hindsight  $\neq$  foresight: the effect of outcome knowledge on judgment under uncertainty. *Quality and Safety in Health Care*. 2003;12(4):304-11.

128. Youngberg B. Event Reporting: The Value of a Nonpunitive Approach. *Clinical Obstetrics & Gynecology* 2008;51(4):647-55.